DOCUMENT RESUME

ED 142 592 95 TM 006 526

AUTHOR	McDonald, Frederick, J.; Elias, Patricia
TITLE	Beginning Teacher Evaluation Study, Phase II, 1973-74. Executive Summary Report.
INSTITUTION	California State Commission for Teacher Preparation and Licensing, Sacramento.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ETS-PR-76-18
PUB DATE	76
NOTE	167p.; For related documents, see ED 127 364-375; Tables may be marginally legible
EDRS PRICE	MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Gains; Classroom Observation Techniques; Data Analysis; Educational Environment; *Effective Teaching; Elementary Education; *Elementary School Teachers; Evaluation Methods; Grade 2; Grade 5; Mathematics; Predictor Variables; Reading; Student Characteristics; Student Testing; *Teacher Behavior; Teacher Characteristics; *Teaching Methods; *Teaching Styles; Test Reliability

ABSTRACT

Significant relationships were found between how teachers teach and how much children learn in reading and mathematics. The study's purpose was to relate classroom observations of experienced elementary school teachers' interactions with their students to changes in the pupils' performance on various measures of reading and mathematical understanding and skill. It was concluded that a pattern of teaching practices is more likely to be related to learning than a single practice, and that effective teaching practices will differ according to subject matter and grade level. In brief, the significance of this study is that effective teaching methods, as well as less effective methods have been identified. This summary describes the methodology and results of the entire study, but does not present the technical details concerning measurement and statistical analysis. Summaries of special studies, intensive studies of the major sources of data and special research studies are included. (MV)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED142592

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PR-76-18

BEGINNING TEACHER EVALUATION STUDY
PHASE II 1973-74

EXECUTIVE SUMMARY REPORT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

BY

FREDERICK J. McDONALD

AND

PATRICIA ELIAS

BEST COPY AVAILABLE

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

BEGINNING TEACHER EVALUATION STUDY, PHASE II

1973-74

EXECUTIVE SUMMARY REPORT

by

Frederick J. McDonald

Patricia Elias

Educational Testing Service
Princeton, New Jersey

A study conducted by Educational Testing Service for the California Commission for Teacher Preparation and Licensing and funded by the National Institute of Education and Educational Testing Service.

Copyright © 1976 by Educational Testing Service. All rights reserved.

The Final Report of Phase II of the Beginning Teacher Evaluation Study has been prepared in five volumes.

Volume I

The Effects of Teaching Performance on Pupil Learning: The Major Results of Phase II of the Beginning Teacher Evaluation Study by Frederick J. McDonald and Patricia J. Elias. (Portions of Chapter 5 were written by Ruth B. Ekstrom and Meredith K. Stone; of Chapter 6 by Barbara Foltin and William Allen). ERIC Document number - ED 127364.

Volume II

Description of the Field Study and Sample by Patricia Wheeler. ERIC Document number - ED 127365.

Volume III

The Observation Systems:

Chapter 1: APPLE Observational Variables and Their Relations to Reading and Mathematics Achievement by Nadine Lambert and Carolyn S. Hartsough. ERIC Document number - ED 127366.

Chapter 2: Reading and Mathematics Observation System: Description and Analysis of Time Expenditures by Robert Calfee and Kathryn Hoover. ERIC Document number - ED 127367.

Chapter 3: The Evaluation of Teaching Behavior Through Observation of Videotape Recordings by Jonathan Sandoval. ERIC Document number - ED 127368.

Volume IV

Pupil and Teacher Tests by Ruth B. Ekstrom. ERIC Document number - ED 127369.

Volume V

Special Studies:

Chapter 1: The Relationship of Teacher Aptitudes to Teaching Behavior by Ruth B. Ekstrom. ERIC Document number - ED 127370.

Chapter 2: Teacher Expectations and Pupil Performance by Marlaine Lockheed. ERIC Document number - ED 127371.

Chapter 3: Historical Data Study by Patricia Wheeler and Patricia J. Elias. ERIC Document number - ED 127372.

Chapter 4: Reading Difficulties: Videotape Test of Teachers' Diagnostic Skills by Patricia J. Elias and Gail Hare. ERIC Document number - ED 127373.

Chapter 5: The Reports of Teachers about Their Mathematics and Reading Instructional Activities by Patricia J. Elias, Gail Hare and Patricia Wheeler. ERIC Document number - ED 127374.

Preface

The study reported here found significant relations between how teachers teach and how much children learn in reading and mathematics. The purpose of the study was to relate observations of experienced elementary school teachers to changes in their pupils' performances on a variety of measures of reading and mathematical understanding and skill. It could have turned out that such differences across a school year were more related to differences in the background of the children. But we found that teaching performances made a significant contribution to such differences and that some teaching practices made a greater difference than others.

These findings should be treated with reasonable caution. They have been developed by observing a relatively small sample of teachers. The results might be different with a new sample. The relations between teaching and learning reported here pertain only to those that were measured in this study. Other characteristics of teaching might be important for other kinds of learning. Those characteristics of teaching found to be more effective in this study *should not be used to evaluate teachers* because the results of the study do *not* describe the differences between competence and incompetence.

If these *caveats* are kept in mind, the results may be used by practitioners to analyze their teaching practices and to develop more effective teaching strategies. Researchers should find the results a source of hypotheses to be tested and a base for theory-building. These results ought to be treated as a first step in an inquiry about the nature of effective teaching.

This study contributes to our understanding of how to help children learn more, understand more, or become more skillful in reading and mathematics. It also describes teaching practices which are less effective.

The effective practices describe a teaching style associated with marked gains on a variety of pupil measures. The less effective or ineffective practices describe a style associated with small gains.

One way to test the validity of results of this study is to train teachers in the effective practices and to study both how they are learned and their effects on pupil learning. Carefully designed experiments of this kind will answer three questions: (1) how "learnable" are these practices; (2) what procedures are effective in helping teacher to learn them; and, (3) do children learn more when teachers have become particularly skillful in these practices?

In brief, the achievement of this study was specifically that it did find teaching effects. It verified that some performances or procedures were associated with greater learning in second- and fifth-grade classes studying reading and mathematics. Some classes learned more than others because their teachers used the effective techniques more frequently than the other teachers did.

This summary describes the methodology and results of the entire study. It does not present all of its technical details. It is intended to give an overview with only that statistical information which will give credibility to the descriptions of the results. Readers interested in the details of the measurement methodology and of the statistical analyses are referred to the Final Report. A brief and non-technical summary is also available.

The full report has been placed in ERIC and the document numbers may be found on page i at the beginning of this summary. Volume I is a description of the design, rationale, results of data analyses and major conclusions. Volume II describes the conduct of the field study in which the data were gathered. Volume III describes the observational systems in detail, and Volume IV the other aspects of the measurement system. Volume V describes the teacher aptitude battery, the study of teacher expectations, the diagnostic film test, the Work Diary instrument, and the study of the data available on classes previously taught by the teachers in this study.

This study was conducted by a team of investigators who worked with the principal investigators and was largely responsible for the conduct of the study: Meredith Stone, William Allen, Barbara Foltin, Ruth Ekstrom, Marlaine Lockheed and David Potter of Educational Testing Service, Princeton; Patricia Elias, Patricia Wheeler, Robert Lambert and Gail Hare of Educational Testing Service, Berkeley; Nadine Lambert and Carolyn Hartsough, University of California, Berkeley; Jonathan Sandoval, University of California, Davis; Robert Calfee and Katherine Hoover, Stanford University. Patricia Wheeler and Robert Lambert were the project's Associate Directors and they, with Patricia Elias, directed and managed the field study.

We wish to express our gratitude to the teachers and administrators who participated in this study. Without them there would have been no study.

We appreciated the opportunity to work closely with the Evaluation of Teaching Committee (as it was then called) of the California Commission for Teacher Preparation and Licensing. This study was conducted for the Commission as a part of a long-range study under their direction. We trust that the work of this phase will be useful to them in designing subsequent phases and in developing policies for teacher preparation and certification.

Frederick J. McDonald

Patricia J. Elias

TABLE OF CONTENTS

Page

Preface iii

INTRODUCTION. 1

THE TEACHERS AND SCHOOLS IN THE STUDY 4

DESIGN OF THE STUDY 5

- The Measurement Strategy 6
- The Measurement of Pupil Characteristics 10
- Observations of Classroom Teaching and Pupil Activities. 14
- The Measurement of Teacher Characteristics 19
- The Measurement of the Characteristics of the Teaching Environment . 23

ANALYSIS OF DATA. 24

THE RESULTS OF THE STUDY. 48

- Grade-Two Reading. 56
- Grade-Five Reading 58
- Grade-Two Mathematic 61
- Grade-Five Mathematics 64
- Suggested Hypotheses 67

SUMMARIES OF SPECIAL STUDIES WITHIN THE TOTAL STUDY 70

INTENSIVE STUDIES OF THE MAJOR PHASE II SOURCES OF DATA 70

- APPLE Observational Variables as Measures of Teacher Performance and Pupil Behavior 70
  - APPLE in the Beginning Teacher Evaluation Study 73
  - Description of Teacher and Pupil Behaviors. 77
  - The Allocation of Classroom Instructional Time. 78
  - Validity of Observational Variables for Predicting Reading Outcomes. 80

TABLE OF CONTENTS (Continued)

Page

Validity of the Observational Data for Predicting Mathematics Outcomes 81

Validity of Observational Variables for Predicting Attitude Outcomes. 82

Time Allocation Data as Predictors of Learning and Attitude Outcomes. 84

Teacher Performances Associated with Different Kinds of Teaching Effectiveness 86

Summary. 90

The RAMOS Observational System 90

Procedures for Conducting a RAMOS Observation. 92

Editing and Analysis of RAMOS Protocols. 94

Results. 95

The Reports of Teachers About Their Instructional Activities. . . . 99

Academic Emphases. 100

The Quality of the Teaching Methodology. 103

The Complexity of the Organizational Structure in the Classroom 103

The Variety and Types of Instructional Materials Used in Reading and Mathematics Programs 105

Teacher Reports of Their Activities and Observations of Those Activities. 106

The Relationship of Teacher Aptitudes to Teaching Behavior. 108

Results. 111

Some Determinants and Consequences of Teacher Expectations Concerning Pupil Performance. 116

Method 117

Results. 118

Summary. 121

TABLE OF CONTENTS (Continued)

Page

PHASE II SPECIAL RESEARCH STUDIES

The Evaluation of Videotape Recordings of Teaching Behavior 121

Results. 123

Summary. 134

Historical Data on Pupil Performance. 134

A Videotape Test of Teachers' Diagnostic Skills 136

Test Development 136

Error Types Selected for Inclusion 137

Test Format. 138

Sample 139

Scoring the Test 140

Results 141

LIST OF TABLES

Table Number — Page

1 Teacher, Student, and School Climate Measures 8

2 Reliability of Grade 2 Pupil Tests. 12

3 Reliability of Grade 5 Pupil Tests. 13

4 Description of Classroom Variables. 18

5 Test Statistics for Fall Teacher Measures 21

6 Test Statistics for Spring Teacher Measures 22

7 Multiple Regression of Residual Scores on Teaching Performances 26

8 Stepwise Regression of Residual Scores on Teaching Performances (Step 3). 28

9 Regression of Student Characteristics and Student Variables on Spring Scores with Fall Scores as the Concomitant Variable. 30

10 Stepwise Regression of Group Scores on Teaching Performance (Step 16). 35

11 Summary of R, R^2 and Percent Variance Accounted for in Stepwise Regressions of Residual, Mean Gain and Pattern Scores on Teaching Performance. 52

12 Examples from APPLE Lexicon for Each Category 72

13 Frequencies of Teacher Activities in Contexts for Teachers with Better than Predicted and Teachers with Poorer than Predicted Outcomes in Second Grade Reading 88

14 Teacher Test Scores 110

15 Teaching Behaviors 112

16 Path Coefficients from Second Grade Teachers' Aptitudes, Cognitive Style, Knowledge, and Attitudes to Their Classroom Behaviors in Teaching Reading and/or Mathematics. . 113

17 Path Coefficients from Fifth Grade Teachers' Aptitudes, Cognitive Style, Knowledge, and Attitudes to Their Classroom Behaviors in Teaching Reading and/or Mathematics. . 114

LIST OF FIGURES

Figure Number		Page
1	A Structural Model of the Domain of Variables Influencing Teaching Performances and Children's Learning . .	7
2	Slope (b) Plotted Against Intercept (c) for Each Teacher	33
3	RAMOS Summary Codes	93

INTRODUCTION

Phase II of the Beginning Teacher Evaluation Study (BTES) was one component in a long-term study of teaching effectiveness sponsored by the California Commission for Teacher Preparation and Licensing. The purpose of this research plan was to develop data on teaching effectiveness and its correlates which will guide the Commission in creating policies on teacher preparation and certification. The first phase was a planning and development stage conducted by the Commission to prepare the research design for Phase II.

The second phase, conducted by Educational Testing Service for the Commission, was concerned with teaching effectiveness in reading and mathematics at the second- and fifth-grade levels and had three purposes:

1. to develop hypotheses about the relations of teaching performances to pupils' learning;
2. to develop a measurement system that might be useful in subsequent phases of the BTES project; and
3. to estimate the relative influence of teachers' characteristics and the environment of teaching on teaching performance and of pupils' characteristics and background on pupils' learning.

One method used to generate hypotheses was, first, to measure selected aspects of pupil learning and to observe teachers in the classroom, and then, to relate statistically the observations of teaching to the measured changes in pupil learning. This method produced significant correlations between the observed teaching performances and the measures of learning. These significant correlations were the basis for generating hypotheses about the relation of teaching performance to pupil learning. These hypotheses are to be tested in subsequent phases of the Beginning Teacher Evaluation Study.

The second purpose of Phase II was to develop or select a variety of measures and to design a measurement system of teaching performances and pupil learning. Tests were developed or selected to assess pupils' decoding and comprehension skills in reading, their ability to apply reading skills to a variety of non-school reading materials, and their attitudes towards reading. We developed or selected tests of pupils' skills in mathematical computations, understanding of mathematical concepts, ability to use concepts and computational skills in solving everyday mathematical problems, and attitudes towards mathematics.

Systems were developed to measure teaching performances by direct observation of classroom teaching actions and activities. These observation systems gathered data on the ways teachers organized children for instruction, on the types and variety of materials they used in instruction, on their interactions with children during instruction, and on the specific content of instruction.

Another important concern of this investigation was to determine the influence of a considerable number of factors on both pupil learning and teaching performances. Measures were developed or selected of the pupils' verbal aptitude, of their expectations for achievement in reading and mathematics, of their cognitive style, and of their background characteristics and out-of-school activities relevant to reading and mathematics. These measures assessed characteristics and experiences which the pupils brought to the classroom. The study investigated the relation of such factors to what the pupils learned in the course of one academic year.

Several features of this measurement strategy should be noted. First, the effects of changes in pupil learning are validly attributed to teachers if relations are found between teaching performances and pupil learning. Effects of previous learning are reflected in the fall scores; the children's socioeconomic backgrounds do not change during the year; and therefore differences among teachers in their pupils' scores are more likely to have been produced by differences in teaching practices.

Second, there is sufficient information in the system to estimate how each measure of a pupil characteristic is related to differences in pupil learning. Third, the interdependencies among the measures may be estimated. We asked, for example, if cognitive style and verbal aptitude were independent? Or were there relations between background factors and pupils' expectations about their own performance? Analysis of such relations enabled us to understand the complex of factors related to pupil learning so that teaching influences on this learning might be more precisely estimated.

It is generally believed that the school environment in which teachers work will influence both what and how they teach. Two important environmental features are the support provided for instructional innovation and experimentation, and the sponsorship and nurturing of creativity and professionalism. How the administration organizes the processes for deciding about the character and substance of the curriculum and the patterns of interpersonal relations in the school may also affect what is taught, how it is taught, and how much the teachers are involved and committed. In Phase II, these aspects of the teaching environment were measured and related to teaching performances.

The measurement system provided data which were used to evaluate four kinds of relations: (1) how teaching actions relate to pupil learning; (2) how pupils' characteristics relate to their learning; (3) how teachers' characteristics relate to how they teach; and, (4) how the environment in which the teacher work affects how they teach. The statistical analyses of these data assessed the relative importance of each area and the degree and direction of the relations of the specific factors to each other.

The Phase II research employed a complex multivariate design. A variety of regression analyses were used to explore relations among variables. The conclusions of the study were developed by comparing and relating the results of many different analyses. These major conclusions are presented in this summary.

THE TEACHERS AND SCHOOLS IN THE STUDY

This study was conducted in eight school districts located throughout the State of California. They included two urban, three suburban, and three rural districts. Forty-three schools provided the sample of teachers and principals. Forty-one second-grade teachers and 54 fifth-grade teachers participated in the study on a voluntary basis. Their teaching experience ranged from 3 to 31 years. About 85 percent of the teachers were white, about 10 percent black, and the rest Chinese-American and Japanese-American. Twenty-five hundred pupils were included in the study.

Since the sample of teachers consisted of volunteers, it cannot be considered a random sample of California teachers. The study did not require a random sample because the goal was not to make general statements about the effectiveness of California teachers. Rather, the goal was to identify teaching performances which, in terms of criteria of pupil learning, were relatively more effective. If the sample included teachers whose teaching practices differ and if those differences were associated with significant changes in pupil learning, the purposes of the study would have been achieved. Differences in pupil learning associated with differences in teaching performances found in this study would not necessarily characterize other effective teachers. The teaching performances also may be a subsample of effective practices. Other research will have to be done to find out both how critical the practices identified in this study are and whether or not there are other effective practices.

DESIGN OF THE STUDY

Data for the study were gathered during the 1973-74 academic year. During this period, pupil achievement levels in reading and mathematics and pupil attitudes towards reading and mathematics were measured twice, once in the fall and again in the spring. Teachers were also tested at these times. During the intervening period, three to eight observations were made of each of the teachers in their classrooms. Approximately 400 days were allotted to observations of classroom teaching. Teachers were also videotaped on one or more days during reading and mathematics instruction. Additional information on the instructional process was collected from the teachers by asking them to complete diaries for two different weeks during the school year. Principals completed questionnaires and were interviewed about their school's programs and teaching environment during the observation period.

The Measurement Strategy

The goal of the measurement strategy was to assess as many factors as feasible that might be related to teaching performances and pupil learning. These factors were conceived of as domains of variables; for example, teachers' aptitudes is a domain of variables. Sixteen different measures of aptitude were used to sample this domain.

Each domain is defined by a generic characteristic of the variables in the domain. Teachers' aptitudes, for example, are variables describing internal perceptual and cognitive processes. The particular measures selected to sample this domain measure different aspects of these processes -- those which might be expected to be related to different ways or styles of teaching.

Seventeen different domains of variables were measured. A model was constructed which represents one way these domains may be related to each other. This model is presented in Figure 1. The core of the model is the three domains of Teaching Performances, Student's Behavior and Learning. Domains of pupil variables are grouped around the Learning domain (on the right-hand side of the diagram) because variables in these domains are likely to have a direct influence on learning. On the left-hand side of the diagram are represented the domains of teacher variables and variables describing the environment of teaching which are expected to have a direct or indirect influence on teaching performances. This model is one theory of how these domains of variables may be related. The validity of the model was tested in this study.

Table 1 lists all the measures used in this study. The measures are listed by domain. The response mode and the time the measure was

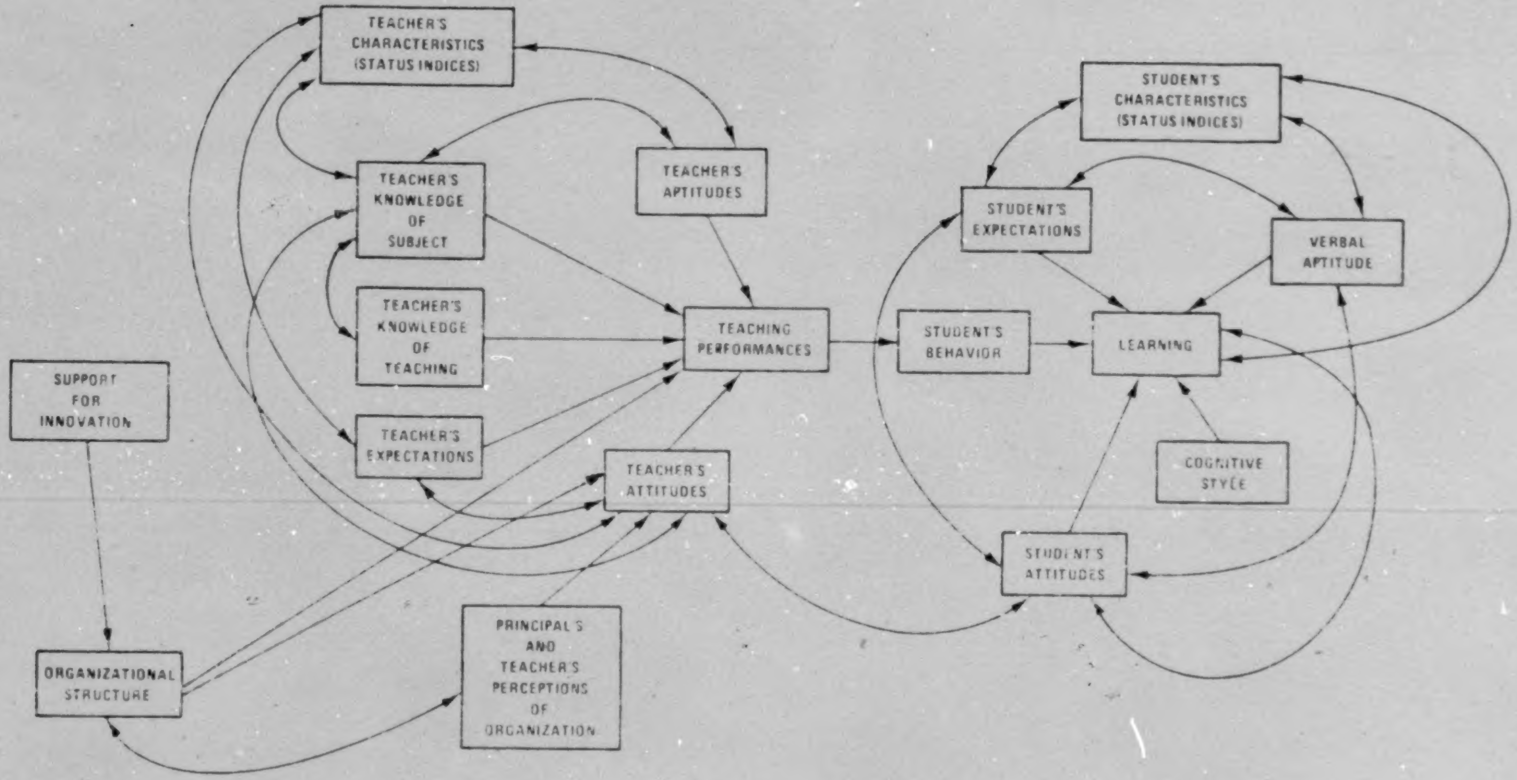

TEACHER'S CHARACTERISTICS (STATUS INDICES)
TEACHER'S APTITUDES
TEACHER'S KNOWLEDGE OF SUBJECT
TEACHER'S KNOWLEDGE OF TEACHING
TEACHER'S EXPECTATIONS
TEACHER'S ATTITUDES
TEACHING PERFORMANCES
SUPPORT FOR INNOVATION
ORGANIZATIONAL STRUCTURE
PRINCIPAL'S AND TEACHER'S PERCEPTIONS OF ORGANIZATION
STUDENT'S CHARACTERISTICS (STATUS INDICES)
STUDENT'S EXPECTATIONS
VERBAL APTITUDE
STUDENT'S BEHAVIOR
LEARNING
COGNITIVE STYLE
STUDENT'S ATTITUDES

Figure 1 A Structural Model of the Domain of Variables Influencing Teaching Performances and Children's Learning

TABLE 1: TEACHER, STUDENT, AND SCHOOL CLIMATE MEASURES

TEACHER VARIABLES	INSTRUMENT	RESPONSE MODES				TIME GIVEN		
		CONSEQUENT	PERFORMANCE	SELF REPORT	OTHER REPORT	BEFORE (FALL)	DURING	AFTER (SPRING)
Knowledge of Subject: Reading Decoding Skills	Letter-sound correspondence subtest	X				X		
	Root Words subtest	X				X		
	Affixes subtest	X				X		
	Syllabification subtest	X				X		
Inferential Reading	Inferential Reading Test	X				X		
Knowledge of Teaching: Reading Knowledge	Primary or Intermediate Knowledge Test	X				X		
Application	Primary or Intermediate Application Test	X				X		
Knowledge of Subject: Mathematics Computation	Arithmetic Computation Test	X				X		
Knowledge of Teaching: Mathematics	General Knowledge of Teaching Mathematics Test	X				X		
Aptitudes	Verbal Facility Test	X				X		
	Vocabulary Test	X				X		
	Memory Test	X				X		
	Figure Classification Test	X				X		
	Deductive Reasoning Test	X				X		
	Inductive Reasoning Test	X				X		
	Mathematics Aptitude Test	X				X		
	Finding Useful Parts	X						X
	Sentence Completion	X						X
	Topics Test	X						X
	Story Surprises	X						X
	Controlled Associations Test	X						X
	Making Sentences	X						X
	Listing Objects	X						X
	Planning Test	X						X
Cognitive Style	Group Embedded Figures Test	X						X
Attitudes	Attitude toward teaching reading			X				X
	Attitude toward teaching mathematics			X				X
Expectations of Students	Rankings in reading and mathematics				X	X		X
Status Characteristics	Teacher Questionnaire			X		X		
Teacher Behavior	APPLE Observations		X				X	
	RAMOS Observations		X				X	
	Videotape Observations		X				X	
	Work Diaries			X			X	

TABLE 1: TEACHER, STUDENT, AND SCHOOL CLIMATE MEASURES
(Continued)

		RESPONSE MODES				TIME GIVEN		
STUDENT VARIABLES	INSTRUMENT	CONSEQUENT	PERFORMANCE	SELF REPORT	OTHER REPORT	BEFORE (FALL)	DURING	AFTER (SPRING)
Learning								
Reading								
Comprehension	CAT Reading Comprehension Test	X				X		X
Achievement	ETS Reading Achievement Test	X				X		X
Application	ETS Reading Application Test	X				X		X
Decoding	ETS Decoding Test	X				X		X
Attitude	ETS Attitude Survey			X		X		X
Mathematics								
Concepts	CAT Mathematics Concepts Test	X				X		X
Computation	CAT Mathematics Computation Test	X				X		X
Application	ETS Mathematics Application Test	X				X		X
Attitude	ETS Attitude Survey			X		X		X
Student Characteristics								
Aptitude	SFTAA	X				X		
Cognitive Style	CPP Group Embedded Figures Test	X						X
Expectations	ETS Pupil Survey			X		X		X
Status Characteristics								
Parent	Parent Questionnaire				X	X		
Teacher	Pupil Survey Instrument				X	X		
Student Behavior	APPLE Observations - Ap-1, 2 and 3		X				X	
SCHOOL CLIMATE VARIABLES								
Teacher	**School Climate Questionnaire**			X				X
	Likert Profile of a School			X				X
Principal	**School Climate Questionnaire**			X			X	
	Likert Profile of a School			X			X	
	Principal Interview			X			X	

taken are indicated. When a Consequent mode is indicated, it means that a test was taken. It was assumed that tests measure the consequences of previous learning. When a Performance mode is indicated, it means that actual behavior in natural situations was observed. These two types of measures were assumed to be direct measures of a variable. Indirect measures are indicated by marks in the Self-Report or Other Report column.

The Measurement of Pupil Characteristics

The California Commission for Teacher Preparation and Licensing specified the categories of pupil learning which should be measured in this study. These were reading (decoding, comprehension, application skills, and attitude towards reading) and mathematics (computational skills, understanding of mathematical concepts, application skills, and attitude towards mathematics). The investigators used the curriculum goals of the State of California to define the specific objectives for each of these subjects. These statements of objectives were reviewed and approved by the Commission. Tests to measure pupils' achievement of these objectives were built specifically for the purposes of this study.

A Decoding Test was constructed to measure such aspects of decoding skills as the recognition of beginning and ending sounds, sight words, and root words. The Reading Comprehension Test measured two aspects of reading comprehension: literal and inferential. The task for the child in literal comprehension was to identify the facts stated in a reading passage; the task in inferential comprehension was to draw an appropriate inference from the information and ideas given in the reading passage. The Reading Application Test presented the child with reading materials not typically found in the reading texts used in schools, such as signs, labels, television schedules, and directions for returning books to the library.

The test contained factual and comprehension questions about these materials. The child was also asked to express his or her feelings about and preferences for reading activities in the Reading Attitude measure.

The Mathematical Computations Test measured a child's skill in adding, subtracting, multiplying, and dividing. A child's understanding of Mathematical Concepts was tested by presenting him or her with problems that could be solved if the child understood a particular mathematical concept. The Mathematical Applications Test required the use of mathematical concepts and computational skill to solve practical problems such as estimating the cost of buying several items on a shopping list. The Mathematics Attitude Test asked the child to express his or her feelings about and preferences for mathematical activities.[1]

In addition to the tests built specifically for the study, the California Achievement Tests in Reading and Mathematics, 1970 Edition, and the Short-Form Test of Academic Aptitude (1970) were also used.[2] Scores on these standardized tests were to be related to scores on the newly developed tests.

An instrument was developed in which children were asked to state their expectations about how well they would do in reading and mathematics that year. A measure of cognitive style[3] that assessed the degree to which the child's characteristic perceptual and cognitive mode was analytic or holistic was also used.

Conventional reliability estimates were calculated on these pupil tests. The alpha coefficients are presented in Tables 2 and 3.

---

[1] The reading, mathematics tests, and other measures used in this study are described in greater detail in Chapter 5 of Volume I and in Volume IV of the Final Report.

[2] Published by CTB/McGraw Hill, Monterey, California.

[3] Forms of the Embedded Figures Test, Consulting Psychologists Press, Palo Alto, California, 1963-71.

Table 2

Reliability of Grade 2 Pupil Tests

Test	Title	Reliability Fall	 Spring
201	CAT Reading Comprehension	.92	.94
202	CAT Math Concepts	.89	.64
203	CAT Math Computation	.95	.86
204a	Attitude toward Reading	.70	.69
204b	Attitude toward Math	.67	.68
204	Attitude Survey - Total	.81	.82
206a	Reading Application - Form 1	--	.84
206b	Reading Application - Form 2	--	.85
206	Reading Application	.84	--
207a	Decoding - Subtests a-g - Form 1	.96	.93
	Decoding - Subtests a-g - Form 2	.96	.93
	Decoding - Subtests a-g - Form 3	.95	.94
	Decoding - Subtests a-g - Form 4	.94	.92
	Decoding - Subtests a-g - Form 5	.95	--
207b	Decoding - Word Sounds - Form 1	.85	.95
	Decoding - Word Sounds - Form 2	.84	.95
	Decoding - Word Sounds - Form 3	.84	.95
	Decoding - Word Sounds - Form 4	.76	.94
	Decoding - Word Sounds - Form 5	.79	--
208	Mathematics Application	.63	.66
209a	Reading Achievement - Form 1	--	.77
209b	Reading Achievement - Form 2	--	.82
209	Reading Achievement	.79	--
210	SFTAA - Total	.94	--
211	Embedded Figures	--	.90

Table 3

Reliability of Grade 5 Pupil Tests

Test	Title	Reliability Fall	Reliability Spring
501	CAT Reading Comprehension	.93	.94
502	CAT Math Concepts	.84	.88
503	CAT Math Computation	.91	.90
504a	Attitude toward Reading	.75	.77
504b	Attitude toward Math	.77	.81
504	Attitude Survey - Total	.80	.82
506a	Reading Application - Form 1	--	.79
506b	Reading Application - Form 2	--	.78
506	Reading Application	.73	--
507a	Decoding - Subtests a-g - Form 1	.95	.92
	Decoding - Subtests a-g - Form 2	.94	.88
	Decoding - Subtests a-g - Form 3	.95	.90
	Decoding - Subtests a-g - Form 4	.95	.92
	Decoding - Subtests a-g - Form 5	.95	--
507b	Decoding - Word Sounds - Form 1	.89	.91
	Decoding - Word Sounds - Form 2	.89	.91
	Decoding - Word Sounds - Form 3	.91	.90
	Decoding - Word Sounds - Form 4	.87	.91
	Decoding - Word Sounds - Form 5	.88	--
508	Mathematics Application	.80	.82
509a	Reading Achievement - Form 1	--	.71
509b	Reading Achievement - Form 2	--	.76
509	Reading Achievement	.73	--
510	SFTAA - Total	.93	--
511	Embedded Figures	--	.88

Information was collected from parents and teachers about a variety of student characteristics including sex and race, socioeconomic background, parents' education, preschool experience, reading at home, watching educational television, absence from school, and the number of schools attended.

The pupils' tests were administered during a two-day session in the fall and again in the spring. The test administrations were conducted by trained substitute teachers, and the investigators provided monitors to facilitate the test administrations. The testing was conducted in each classroom. The tests had been pilot-tested to assure their suitability for second- and fifth-grade pupils.

Observations of Classroom Teaching and Pupil Activities

The goal of the procedures used in the observations of classroom teaching was to obtain as complete a record as possible of what the teachers did and how they interacted with the children during the observation periods. Two systems of observation were used in each classroom. One system, APPLE (Anecdotal Process for Promoting the Learning Experience),[4] focused primarily on the individual pupil and his or her interactions with the teacher, and required an observer to write down behavioral statements of what the teacher and pupils were doing. These statements were subsequently categorized by independent coders using a lexicon of behavioral descriptions.

The other system, RAMOS (Reading and Mathematics Observation System),[5] was developed originally by intensively observing reading teachers and was extended to cover observations of mathematics instruction. During its development detailed behavioral recordings had been made of what teachers did during reading periods. (The RAMOS system focused primarily on the teacher.) These observations were subsequently organized into a system of categories and codes for each category. The category system developed in this way was the one used during the observation periods of this study.

---

[4] Developed by Nadine Lambert, University of California, Berkeley, 1970.

[5] Adaptation of Reading Diary, by Robert Calfee and Kathryn Hoover, Stanford University, 1972.

A preselected sample of eight pupils in each class was observed during each observation. This sample was selected to represent the range of differences among the children in terms of their teacher's expectations about their performance compared to their measured achievement. The ethnic composition of the class was preserved in the sample, and boys and girls were equally represented.

The observational data were used to create larger categories of teacher and pupil behavior for which scores were generated for each classroom. The larger categories described the teaching contexts, the instructional activities, the variety of instructional goals and materials, and the techniques of instruction. Each such category contained subcategories which described the variations in teaching activities and actions observed. For example, the category teaching context described how the teacher organized the class for instruction -- in groups, with a class-as-a-whole, or with each child working independently. Teaching activities described how instruction was given -- by questioning, explaining, discussing, or monitoring independent reading or the working of problems.

Two aspects of these observation systems and the data they produced are worth noting. First, the systems give a comprehensive picture of classroom activity on the occasion of observation. Prior judgments had not been made about what features of this activity might be more important than others. Second, when the observational data were analyzed, it was found that: (1) some events occurred relatively infrequently and the relative proportions of the more frequent events were not large; and, (2) the categories were intercorrelated in varying degrees. This second feature of the data suggested that teachers differed in the profiles of their scores. The stepwise regression methods used in the analysis to estimate the relation of the teachers' performance scores to pupil scores take into account this intercorrelation among the category scores.

Videotapes provided another way to collect information on what was happening in the classrooms during reading and mathematics instruction. All teachers were videotaped on one to four occasions. The data from the subset of teachers who were videotaped more than once were used to study the consistency of teaching performances on different occasions.

It had been planned to use APPLE and RAMOS categories to describe the teaching performances portrayed on the videotapes. But the videocamera did not pick up all the events that could be seen and heard during live observations. A new system was therefore developed which reflected the recorded events; but some APPLE and RAMOS categories were adopted as part of this system.[6]

The data from these recorded observations were analyzed separately because even though live and videorecorded observations had been made on the same occasions the resulting videorecords were obviously a small sample of all the events that had been recorded during the live observations. We concluded that videorecordings had a limited usefulness for studying teacher performance unless two or more cameras were used -- a system which would probably change the nature of the situation.

---

[6] Developed by Jonathan Sandoval, University of California, Davis, 1974. These data were analyzed separately. The results of the analysis of the videotapes appear in Chapter 3, Volume III of the Final Report.

Another source of information on the teachers' activities was an indirect measure of teaching performance. The teachers provided a diary of their activities for two separate weeks of teaching.[7] This diary was developed in consultation with the teachers. In the diaries the teachers described the time spent in direct instructional activities, the range of content covered during instruction, the complexity of the instructional system, and the variety of materials used.

The three major sources of data -- the two direct observational systems and the diary -- provided the scores for each teacher which described the teacher's activities and actions. Originally these sources provided 136 scores for each teacher. This number was reduced by grouping and an appropriate combinatorial method to 22 scores for each teacher. The categories of teaching performance and pupil behaviors used in the major analyses are presented in Table 4.[8]

Each teacher was observed with each system and completed the Work Diary, producing 22 scores. As previously mentioned, some of the scores are

---

[7] The description of the Work Diary instrument and the analyses made of the data gathered with it are presented in Chapter 5, Volume V of the Final Report.

[8] The categories of each observational system were separately related to measures of pupil learning by correlating each category score with pupil spring scores partialling out their fall scores. These analyses as well as frequency data were used to construct the superordinate system of 22 categories listed in Table 4. The results of these basic analyses are contained in Volume III, Chapters 1 and 2, and Volume V, Chapter 5. It eas necessary to combine the categories to reduce their number because there were more categories than teachers. The method of regrouping the original categories is described in Volume I, Chapter 4.

Table 4

DESCRIPTION OF CLASSROOM VARIABLES

Teaching Performance Variables

WD-1: An index of direct teaching time in the classroom.

WD-2: An index of the range of content or skills taught.

WD-3: Quality of a teacher's conceptualization of instructional activities.

WD-4: An index of the variety of instructional contexts or settings in a classroom (weighted by teacher doing most of the teaching).

WD-5: An index of the variety of instructional materials used.

R-1: Roles: An index of the variety of instructional roles assumed by a teacher (weighted by time spent in direct instruction and facilitation).

R-2: Aims: An index of the variety of instructional goals (weighted by practice and review).

R-3: Materials: An index of the variety of instructional materials (books, workbooks, paper and pencils).

R-4: Activities: An index of the variety of instructional activities (weighted by individual seatwork).

AP-5: Teaching context: An index of the time pupils spend working independently.

AP-6: Teaching context: An index of the time pupils spend in group activities with the teacher.

AP-7: Teaching context: An index of the time pupils spend in teacher-led activities that involve the whole class.

AP-8: Teaching activity: Organizing and maintaining the instructional process (weighted by work with an individual, group or class).

AP-9: Teaching activity: Monitoring pupil progress (weighted by short term interactions such as answering, asking, checking).

AP-10: Teaching activity: Developing pupil understanding (weighted by sustained interactions such as explaining, discussing, questioning).

AP-11: Teaching activity: Teacher mobility (weighted by circulating).

AP-12: Teaching response: Responsiveness to individual pupils (weighted by positive feedback, praise).

AP-13: Teaching response: Control of individual pupil behavior (weighted by redirecting).

AP-14: Teaching response: Nonresponsiveness to individual pupil behavior.

Pupil Behavior Variables

AP-1: Pupil demonstration of knowledge or understanding of content during instructional activities.

AP-2: Pupil inattentive or off-task behaviors during instructional activities

AP-3: Pupil attentive or on-task behaviors during instructional activities

Copyright © 1976, Beginning Teacher Evaluation Study.
Educational Testing Service, Princeton, New Jersey

measures of similar occurrences. RAMOS-3, for example, is an index of the variety of materials used by pupils during instructional activities as directly observed; Work Diary-5 also measures the variety of materials used but as reported by the teacher. The 22 scores are presumed to be a sample of a teacher's teaching style.

When the original scores were reduced to the 22 categories, some of these scores were measures of similar but not necessarily identical teaching activities and performances. The 22 categories, however, could be grouped as follows (R = RAMOS score; AP = APPLE score; WD = Work Diary score):

1. Instructional Time: WD-1, R-1.
2. Instructional Content: WD-2, WD-3, R-2.
3. Instructional Material: WD-5, R-3.
4. Instructional Organization: WD-4, AP-5, AP-6, AP-7.
5. Instructional Activity: R-4, AP-8, AP-9, AP-10, AP-11, AP-12.
6. Social Control Interactions: AP-13, AP-14.
7. Pupil Behavior: AP-1, AP-2, AP-3.

Teachers' scores differed across these measures. Some teachers, for example, did not teach the class-as-a-whole; others did. Some used a variety of instructional materials; others did not. Teachers differed in how they interacted with pupils. It was such differences in the patterns of performance among the teachers that were related to differences in pupils' learning.

The Measurement of Teacher Characteristics

Teachers' knowledge, aptitudes, and attitudes were also measured. The tests of knowledge were of two kinds: knowledge of the subject taught (reading and mathematics) and knowledge of teaching methodology. Two attitude measures were used, one pertaining to attitudes toward the teaching of reading, the other to attitudes toward the teaching of mathematics. The teacher-aptitude battery included measures of verbal and mathematical aptitude, organization of information, reasoning, and originality; it also contained a measure of cognitive style which, like the pupil measure, assessed whether the teacher's perceptual and cognitive style was more analytical or holistic in character. A portion of the teacher-aptitude battery was given in the fall and the remainder of the battery in the spring.

Teachers' expectations for their pupils' achievement in reading and mathematics were measured by asking the teachers to rank-order each pupil in terms of predicted standing in reading and in mathematics at the end of the school year. Information was collected from the teachers on their own family background, education and teaching experience, their aspirations for and satisfactions with teaching, and their preferences for teaching pupils with different kinds of home, community, and ability characteristics.

These instruments were used to investigate many kinds of relations between teacher characteristics and teacher performances in the classroom. Presumably such stable characteristics as aptitudes influence how a teacher perceives, organizes, and interacts with pupils while teaching. Similarly, teachers' expectations for how pupils will learn and what they will accomplish might influence how they teach each pupil and the class. Teachers' attitudes towards reading and mathematics and their knowledge of these subjects presumably also influence how they teach.

The teacher-aptitude tests were factor analyzed, and the analysis yielded four factor scores: flexibility, verbal fluency, reasoning, and memory. In all, ten scores were derived to describe teachers' attitudes, aptitudes, and knowledge. The reliabilities of the original aptitude measures, the factor scores, and of the other teacher measures are presented in Tables 5 and 6.

TABLE 5

Test Statistics for Fall Teacher Measures

Test Title	Number of Teachers	Number of Items	Total Group Scores-Mean	Total Group Scores-S.D.	Reliability
VERBAL FACILITY	97	30	24.94	3.36	.72
VOCABULARY	97	18	11.64	3.05	.77
DECODING	97	300	281.84	17.95	**
ROOT WORDS	97	45	44.87	0.53	.54
AFFIXES	97	15	14.97	0.17	**
SYLLABIFICATION	97	30	27.24	1.69	.50
INFERENCE	97	10	6.83	2.00	.59
PRIMARY KNOWLEDGE	42	6	4.57	0.79	*
PRIMARY APPLICATION	42	35	14.88	3.23	.53
INTERMEDIATE KNOWLEDGE	55	7	5.00	1.28	*
INTERMEDIATE APPLICATION	55	35	13.62	3.42	.67
MEMORY	97	21	13.06	4.86	.84
PICTURE GROUPING	97	112	52.92	15.80	.48***
NONSENSE SAYINGS	97	15	6.88	2.77	.67
INDUCTIVE REASONING	97	15	10.48	2.57	.71
MATHEMATICS COMPUTATION	97	80	32.64	7.90	.79***
MATHEMATICS APTITUDE	97	21	7.75	3.55	.74
TEACHING MATH	97	16	8.89	2.49	.51

* = Reliabilities not computed due to short length of these tests.

** = Reliability not computed due to lack of variance on this measure.

*** = Communalities which represent lower boundary reliability estimates.

TABLE 6

Test Statistics for Spring Teacher Measures

Test Title	# of Teachers	# of Items	Mean	Standard Deviation	Communalities* Grade 2	Communalities* Grade 5
Finding Useful Parts	94	10	6.98	2.07	.70	.63
Embedded Figures	94	25	17.98	4.95	.66	.71
Sentence Completion	94	70	26.90	12.00	.75	.72
Topics	94	**	27.52	6.19	.67	.76
Story Surprises	94	6	3.49	2.16	.70	.64
Controlled Associations	90	8	41.86	13.60	.67	.54
Making Sentences	94	14	11.03	2.59	.86	.69
Listing Objects	94	**	25.60	5.50	.74	.68
Planning	94	**	7.28	1.92	.84	.64
Attitude - Teaching Reading		10	8.67	1.54	.82	.82
Attitude - Teaching Mathematics	94	10	8.96	1.81	.74	.77

* Communalities represent a lower-boundary reliability estimate.

** Free response items in test; maximum score not known.

The Measurement of the Characteristics of the Teaching Environment

Information was gathered from the principals and teachers on characteristics of each school and on the teacher-principal relations as perceived by both teachers and principals. Each teacher and principal completed a scale measuring characteristics of the organizational climate.[9] The teachers also responded to a questionnaire about their participation in curriculum decisions and decisions pertaining to instructional methods.[10]

Each principal responded to a questionnaire about the school. The principals were asked to describe the organization of the staff, the programs and goals of instruction, teaching methods and materials, and the district's support for innovation and experimentation. Each principal was also interviewed about the processes for making instructional and curricular decisions in the school and his or her goals and problems.[11]

This information was organized into categories which described support for innovation, organizational structure, principals' perceptions of organizational climate and teachers' perceptions of organizational climate. The scores were used as the measures of each teacher's environment of teaching.

---

[9] Teacher and Principal Forms of Profile of a School, Rensis Likert and Associates, Inc., Ann Arbor, Michigan, 1971.

[10] Adaptation of form developed by the staff of the Environments for Teaching Program, directed by Elizabeth G. Cohen, Stanford Center for Research and Development in Teaching, Stanford, California, (NE-C-00-3-0062).

[11] Forms for this interview were developed by the staff of the Environments for Teaching Program, directed by Elizabeth G. Cohen, Stanford Center for Research and Development in Teaching, Stanford, California, (NE-C-00-3-0062).

ANALYSIS OF DATA

Four major kinds of analyses were performed on the scores derived from the data collected. The purposes of the analyses were first, to determine whether the effects of teaching were significant -- did teachers make an appreciable difference in how much pupils learned; second, to isolate those aspects of teaching producing these effects -- which teaching activities accounted for greater changes in pupils' learning; and third, to find the correlates of the differences among the teachers in teaching performance.

These questions were addressed by performing multiple and stepwise regression analyses of two kinds of pupil scores -- mean difference and residual scores -- on the measures of teaching performances. These types of analyses yield multiple correlation coefficients which are a measure of the degree of relation between the teaching performances and the pupil scores. When significant, the multiple correlation coefficient is an estimate of the amount of variance in the scores that is associated with differences in teaching performances.

The stepwise regression analyses are one kind of multiple regression in which all of the teaching-performance variables are used in proportion to their power to predict differences in pupils' scores. The teaching variables were entered into the analysis in the order of the magnitude of their contribution to the multiple correlation coefficient. The stepwise regression yields a multiple correlation coefficient and estimates the positive or negative contribution of each variable. The result is a pattern of variables which predicts differences in pupils' scores.

Two kinds of scores were used to measure differences in achievement -- residual gain scores and mean difference scores. The former score is a measure of the change, with the contribution of the initial (fall) score controlled for or partialled out. (The fall scores were highly correlated with the spring scores.) When the residual scores are regressed against the teaching-performance scores, the result is an estimate of how much the teaching-performance scores are associated with the spring scores, over and above the correlation between fall and spring scores. The mean change score, the second type of score used, is the mean or average of all the differences in individual pupil fall and spring scores.

Some teaching performances may be related to one kind of score but not the other. Some may be related to both. The methods of analysis provide ways of estimating the different effects of the teaching performances. Any teaching performance variable that correlated significantly with both of these scores is likely to have a strong enough relation to pupil performance to be considered in future studies.

The multiple and stepwise regression methods are ways of estimating the predictive power of a group of variables. For Grade 2 Reading, for example, we used the Spring Total Reading Score as the criterion measure or dependent variable. The predictor variables are the teaching performance variables (22 such variables) and the Fall Total Reading Score. Table 7 presents the results of the multiple regression analysis for Grade 2 Reading.

The first item of data to note is the magnitude of R and its significance. Here $R = .86$ and $p = .44$. In this case, as large as R is, it is not statistically significant. Therefore, the teaching performance variables are unreliable predictors of the variance in Spring Total Reading Scores.

Table 7

Multiple Regression of Residual Scores on Teaching Performances
Grade 2 Reading (N = 33)

```
GAIN ANALYSIS   GRADE 2 READING                                   E

THE DEPENDENT VARIABLE IS TCTREADS.  THE MULTIPLE CORRELATION IS  0.8578.  THE STANDARD ERROR OF ESTIMATE=          0.0698

                                            PROPORTION                                                          PROBABILITY
                             SUM OF SQUARES  OF SQUARES        N.D.F.   MEAN SQUARE        F RATIO        OF LARGER F
   TOTAL ABOUT ORIGIN                1.0000                       32.
UNDER NULL HYPOTHESIS                0.1660      1.0000           31.
    DUE TO HYPOTHESIS                0.1221      0.7359           22.        0.0056         1.1399           0.4408
                ERROR                0.0438      0.2641            9.        0.0049

                 STANDARD      REGRESSION     STANDARD      T STATISTICS   CONTRIBUTION   MEASURE OF
                REG. WEIGHT     WEIGHTS      ERROR OF WT.  WITH    9.D.F.    TO R-SQ.     COLLINEARITY
CONCOMITANT VARIABLES

    TOTREADF                     0.9901        0.1293        7.6599                          0.7085

INDEPENDENT VARIABLES

      RD-1       -1.0812        -0.4408        0.2753       -1.6013        -0.0752          0.9356
      RD-2       -0.2743        -0.1118        0.1740       -0.6424        -0.0121          0.8390
      RD-3       -0.2009        -0.0819        0.1300       -0.6296        -0.0116          0.7119
      RD-4       -0.3711        -0.1514        0.1375       -1.1005        -0.0355          0.7419
      RD-5        0.5909         0.2503        0.1875        1.3354        -0.0523          0.8501
      RAMOS-1     0.1715         0.0701        0.2093        0.3351        -0.0033          0.8880
      RAMOS-2    -0.6993        -0.2852        0.2510       -1.1362        -0.0579          0.9225
      RAMOS-3     0.9547         0.4005        0.1616        2.4782         0.1802          0.8023
      RAMOS-4     0.0327         0.0134        0.1486        0.0905        -0.0002          0.7747
      APPLE-1     0.3795         0.1546        0.1425        1.0852        -0.0346          0.7600
      APPLE-2    -0.0261        -0.0106        0.1471       -0.0723        -0.0002          0.7744
      APPLE-3     0.2439         0.0995        0.2790        0.3530        -0.0038          0.9368
      APPLE-5     0.1661         0.0679        0.1275        0.5326        -0.0093          0.6982
      APPLE-6    -0.5191        -0.2181        0.2284       -0.9549        -0.0268          0.9007
      APPLE-7     0.1379         0.0631        0.2005        0.3147        -0.0029          0.6473
      APPLE-8    -0.1437        -0.0611        0.1225       -0.4983        -0.0073          0.6479
      APPLE-9    -0.3860        -0.1573        0.2596        0.6057        -0.0108          0.9278
      APPLE-10    0.5777         0.2335        0.2175        1.0728        -0.0344          0.8969
      APPLE-11    0.2379         0.1008        0.1361        0.7407        -0.0161          0.7154
      APPLE-12    0.5824        -0.2373        0.2049       -1.1581        -0.0394          0.8840
      APPLE-13    0.0135         0.0057        0.2053        0.0276        -0.0000          0.8776
      APPLE-14   -0.2157        -0.0900        0.1471       -0.6118        -0.0110          0.7639
```

The next step was to perform a stepwise regression, a procedure which uses the predictor variables successively in terms of their contribution to R^2. The results of this analysis for Grade 2 Reading are presented in Table 8.

$R = .69$ and $p = .0005$. By computing R^2 and changing it to a percentage by multiplying by 100 percent, we find that the three predictor variables, R-3, R-2, and WD-1, predict 46 percent of the variance after fall scores are partialled out. In this analysis fall scores are the major contributors to the prediction (read across from TOTREADF).

With these data we can write a prediction equation:

$$X_S = .96X_F + .27X_{R-3} - .18X_{R-2} - .14X_{WD-1}$$

We are reasonably certain that predictions of spring scores made by using this equation will be accurate.

The methods described above are ways of finding out which combination of predictor variables predict best the variance in the Spring Total Reading Scores. The conclusions from the analyses presented in Tables 7 and 8 are: (1) the teaching performance variables taken as a set do not add to the power of the prediction that can be made from the fall score along (see Table 7); (2) if the contribution of each variable to R^2 is considered, however, we find that three teaching performance variables predict 46 percent of the remaining variance after fall scores are partialled out (see Table 8).

The fall scores in all analyses were the best predictors of spring scores. They predicted 64 percent to 81 percent of the variance in spring scores. Hence, there is 36 to 19 percent of the remaining variance to be predicted. The addition of three teaching performance variables predicts almost half (46 percent) of this remaining variance. Thus we can make a

Table 8

Stepwise Regression of Residual Scores on Teaching Performances (Step 3)
Grade 2 Reading (N = 33)

```
THE DEPENDENT VARIABLE IS TOTREADS.  THE MULTIPLE CORRELATION IS  0.6784.  THE STANDARD ERROR OF ESTIMATE=         0.0566

                                            PROPORTION                                                    PROBABILITY
                          SUM OF SQUARES    OF SQUARES        N.D.F.   MEAN SQUARE       F RATIO       OF LARGER F
      TOTAL ABOUT ORIGIN      1.0000                            32.
   UNDER NULL HYPOTHESIS      0.1660          1.0000            31.
      DUE TO HYPOTHESIS       0.0764          0.4602             3.       0.0255          7.9569          0.0005
                  ERROR       0.0896          0.5398            28.       0.0032

                  STANDARD      REGRESSION     STANDARD      T STATISTICS    CONTRIBUTION    MEASURE OF
                 REG. WEIGHT     WEIGHTS     ERROR OF WT.  WITH   28.D.F.     TO R-SQ.     COLLINEARITY
   CONCOMITANT VARIABLES

      TOTREADF                    0.9641        0.0586        16.4539                         0.0681

   INDEPENDENT VARIABLES

      RAMOS 3       0.6438        0.2701        0.0666         3.9369        -0.2988          0.2791
      RAMOS-2      -0.4482       -0.1828        0.0667        -2.7387        -0.1446          0.2802
      WD 1          0.3451       -0.1407        0.0568        -2.4770        -0.1133          0.0068

   OTHER VARIABLES                                        NOW    27.D.F.

      WD 4          0.1816       -0.0741        0.0613        -1.2082         0.0277          0.1607
      WD-5          0.1744        0.0824        0.0602         1.3666         0.0350          0.0736
      RAMOS-1      -0.0383       -0.0157        0.0735        -0.2134         0.0009          0.3817
      WD 2          0.0293        0.0119        0.0621         0.1922         0.0007          0.1412
      WD-3          0.0010        0.0004        0.0561         0.0071         0.0000          0.0165
      RAMOS-4      -0.0715       -0.0295        0.0752        -0.3919         0.0031          0.4034
      APPLE-1       0.0570        0.0232        0.0601         0.3862         0.0030          0.0370
      APPLE-2      -0.0761       -0.0310        0.0595        -0.5214         0.0054          0.0705
      APPLE-3      -0.1676       -0.0684        0.0583        -1.1740         0.0262          0.0670
      APPLE-5       0.0769        0.0314        0.0578         0.5439         0.0059          0.0101
      APPLE-6      -0.0421       -0.0177        0.0533        -0.2574         0.0013          0.2556
      APPLE-7       0.0411        0.0168        0.0663         0.2636         0.0016          0.0507
      APPLE-8      -0.2055       -0.0473        0.0594        -1.2706         0.0307          0.2721
      APPLE-9      -0.0434       -0.0177        0.0607        -0.2914         0.0017          0.1030
      APPLE-10      0.0227        0.0092        0.0591         0.1556         0.0005          0.0488
      APPLE-11     -0.0005       -0.0002        0.0678        -0.0034         0.0000          0.2180
      APPLE-12     -0.1016       -0.0414        0.0696        -0.5948         0.0070          0.3231
      APPLE-13     -0.0618       -0.0259        0.0601        -0.4317         0.0037          0.0323
      APPLE-14     -0.0947       -0.0395        0.0600        -0.6532         0.0085          0.0491
```

fairly accurate prediction of a pupil's spring reading score by knowing his or her fall score and three pieces of information about how his or her teacher teaches.

A relevant question in prediction studies is, is there another set of predictors which predict spring scores as well as or better than teaching performance variables? An obvious set for such consideration are those describing the background characteristics of pupils. Table 9 presents the results of a multiple regression analysis using 28 background and aptitude characteristics. Here $R = .49$, $p < .001$, and the set of predictors account for 24 percent of the variance after fall scores are partialled out.

The multiple R produced in this analysis ($R = .49$) is smaller than that produced in the multiple regression of residual scores on teaching performance variables ($R = .86$, see Table 7) but is significant whereas the R using the teaching performance variables is not.

A stepwise regression of pupils' characteristics was run, entering first pupils' expectations for achievement as measured in the spring (S-EXP--S). This variable had the largest contribution to R^2 in the multiple regression analysis (see column labelled, "Contribution to R-SQ," opposite "S-EXP--S" in Table 9). Its addition predicted nine percent of the remaining variance. When SES_1, the variable with the next largest contribution to R^2, was added, 11 percent of the remaining variance was predicted by these two variables. Additional entries contributed very little to the prediction.

It is reasonable to conclude from these analyses that adding teaching performances to the prediction gives a better prediction than adding pupil characteristics variables. This pattern was found both for reading and mathematics and at both grades.

Table 9

Regression of Student Characteristics and Student Variables on Spring Scores with Fall Scores as the Concomitant Variable; Grade 2 Reading

THE DEPENDENT VARIABLE IS TOT-RD-S, THE MULTIPLE CORRELATION IS 0.4878, THE STANDARD ERROR OF ESTIMATE= 0.0237

	SUM OF SQUARES	PROPORTION OF SQUARES	N.D.F.	MEAN SQUARE	F RATIO	PROBABILITY OF LARGER F
TOTAL ABOUT ORIGIN	1.0000		500.			
UNDER NULL HYPOTHESIS	0.3468	1.0000	499.			
DUE TO HYPOTHESIS	0.0825	0.2379	28.	0.0029	5.2521	0.0000
ERROR	0.2643	0.7621	471.	0.0006		

	STANDARD REG. WEIGHT	REGRESSION WEIGHTS	STANDARD ERROR OF WT.	T STATISTICS WITH 471.D.F.	CONTRIBUTION TO R-SQ.	MEASURE OF COLLINEARITY
CONCOMITANT VARIABLES						
TOT-RD-F		0.5450	0.0346	15.7475		0.5315
INDEPENDENT VARIABLES						
SEX---PS	0.0993	0.0587	0.0250	2.3450	-0.0089	0.0977
RACE--PS	-0.0470	-0.0278	0.0257	-1.0792	-0.0019	0.1484
SES---PS	-0.0160	-0.0105	0.0355	-0.2951	-0.0001	0.4492
SES---PQ	0.0283	0.0177	0.0338	0.5223	-0.0004	0.4471
M-ED--PQ	0.0970	0.0633	0.0404	1.5665	-0.0040	0.5776
F-ED--PQ	0.0736	0.0477	0.0453	1.0572	-0.0018	0.6661
MMATH-PQ	-0.1479	-0.0916	0.0324	-2.8248	-0.0129	0.4094
FMATH-PQ	0.0743	0.0465	0.0340	1.3686	-0.0030	0.4506
P-ACT-PQ	0.0333	0.0197	0.0248	0.7950	-0.0010	0.0798
FIHPL-PQ	0.0164	0.0097	0.0257	0.3773	-0.0002	0.1433
HEADS-PQ	-0.0386	-0.0229	0.0264	-0.8663	-0.0012	0.1367
TITL1-PS	0.0145	0.0086	0.0243	0.3520	-0.0002	0.0469
BILIN-PS	-0.0593	-0.0351	0.0248	-1.4130	-0.0032	0.0907
PRMIG-PS	-0.0940	-0.0562	0.0259	-2.1729	-0.0076	0.1366
MILIT-PS	-0.1638	-0.0983	0.0260	-3.7861	-0.0232	0.1351
OTHER-PS	-0.0436	-0.0267	0.0255	-1.0464	-0.0016	0.1362
PRESC-PQ	0.0316	0.0192	0.0270	0.7133	-0.0008	0.1747
EL.CL-PQ	-0.0143	-0.0084	0.0245	-0.3434	-0.0002	0.0636
C-ECA-PQ	0.0517	0.0308	0.0258	1.1953	-0.0023	0.1336
F-LT-ABS	0.0351	0.0211	0.0270	0.7809	-0.0010	0.1978
ABSI-ABS	-0.0714	-0.0428	0.0257	-1.6663	-0.0045	0.1195
S-ATT-PS	-0.0355	-0.0210	0.0244	-0.8611	-0.0012	0.0476
S-EXP--F	0.0429	0.0312	0.0278	1.1231	-0.0020	0.1454
S-EXP--S	0.1461	0.0946	0.0285	3.3211	-0.0178	0.1534
S-ATT--F	-0.0121	-0.0072	0.0267	-0.2686	-0.0001	0.2036
S-ATT--S	0.0472	0.0287	0.0281	1.0230	-0.0017	0.2714
SETAA	0.2013	0.1479	0.0353	4.1944	-0.0285	0.2975
COGSTYLE	0.1240	0.0801	0.0289	2.7739	-0.0124	0.1908

The first step in the analyses was to perform these multivariate regressions with both residual and mean difference scores. The purpose was to find teaching performance variables which appeared regularly in the different analyses and to find the variables which predicted one kind of score but not the other. The goal was to find the best set of predictors for each criterion (pupils' scores in reading and mathematics).

These methods assume, however, the performance of each teacher's class is similar in one respect, the rate of change from fall to spring. This pattern of change can be studied directly by computing the regression of spring scores on fall scores for each class. The slope of the regression line estimates this rate of change.

A third type of analysis was used to determine whether there were systematic differences in the patterns of change and if teaching performances were associated with these patterns. A regression analysis of the spring scores on fall scores was performed for each class. This analysis showed that the classes differed systematically in the patterns of differences between fall and spring. Some classes were characterized by substantial gains which were spread more or less evenly across the class; that is, a pupil's fall score was not a strong predictor of his or her final score in the spring. Other classes were characterized by relatively small gains, but the greatest gains were made by pupils whose fall scores were high; that is, fall scores accurately predicted the amount of gain of an individual pupil -- lower fall scores were associated with smaller gains, and higher fall scores with larger gains.

These patterns appeared at both grade levels and in both reading and mathematics. Multiple and stepwise regressions were performed with scores that defined these pattern differences and with teaching-performance scores to determine if teaching performances were associated with the differences in patterns.

An example of these differences in the patterns of gains is provided for Grade 2 Reading in Figure 2. This figure was conceptualized as follows: (1) spring total reading scores were regressed against total fall scores for each class separately; (2) these data revealed that the correlations between fall and spring scores were different for different teachers; (3) plots of the regression line for each class showed considerable differences in the slopes of the regression lines; and (4) these differences were summarized by constructing Figure 2.

The horizontal axis in Figure 2 represents the range in the values of the slopes of the regression lines; the vertical axis represents the range in the values of the intercept of the line (i.e., the point where the regression line would intersect the spring-score axis). The spring and fall mean scores of each teacher's class are represented by a point in Figure 2. The regression line for T86, for example, is represented by a slope of .60 and an intercept of 79.25; T11, by a slope of 1.21 and an intercept of -31.63. (Note: a negative intercept does not mean that the class did not gain in measured reading skill; it is an indication of how steeply sloped the regression line is.)

This array indicated the differences in the pattern of gains. In T86's class all pupils made gains; in T11's class, large gains were made by pupils with high fall scores; pupils with low fall scores made negligible gains.

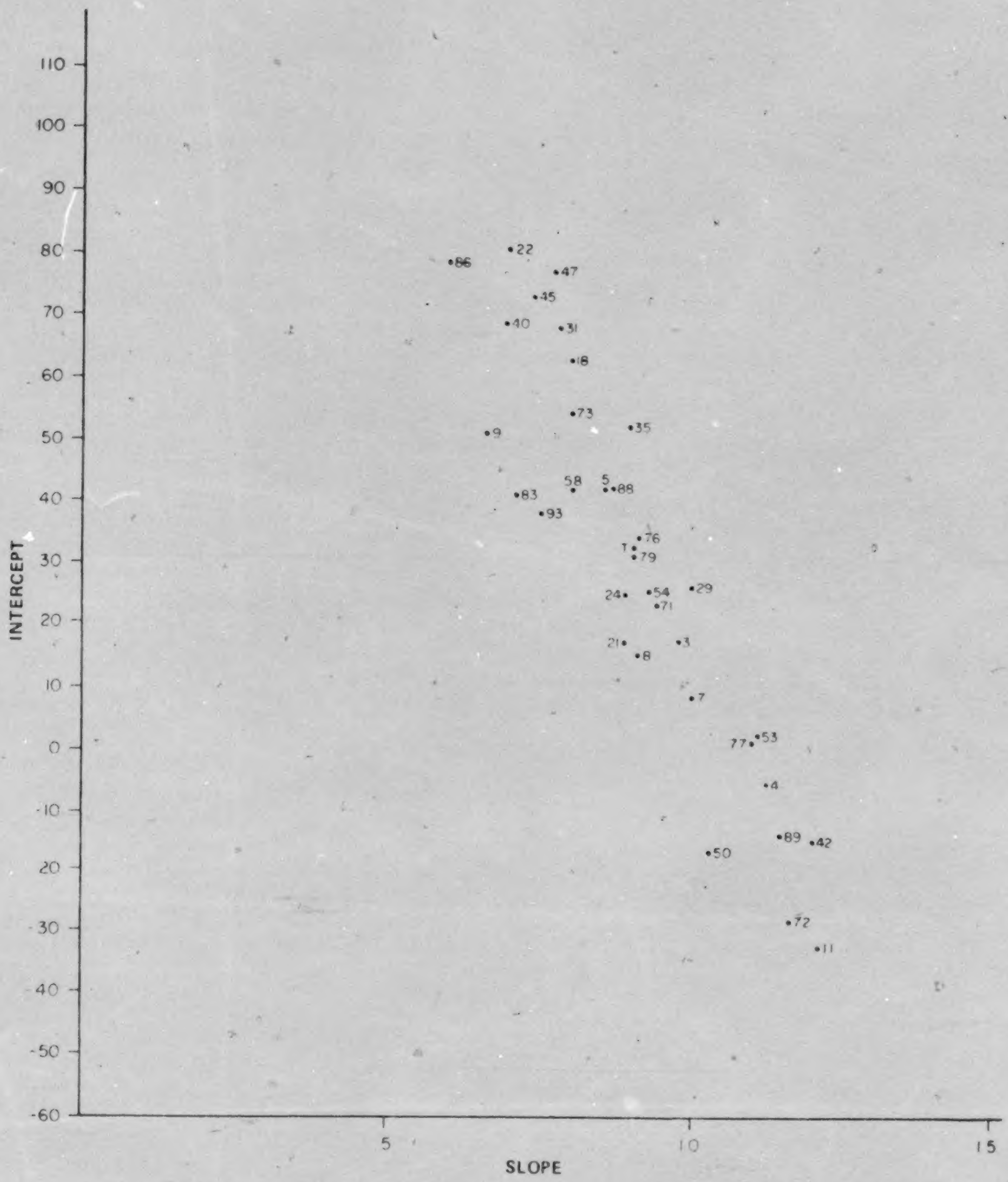


Figure 2. Slope (b) Plotted Against Intercept (c) For Each Teacher; Grade 2 Reading (N = 33)

These differences were so striking that we assigned scores to represent them. Then these new scores were regressed against teaching performance scores. This analysis told us if teaching performances accounted for the differences in patterns of gains. Table 10 presents the results of the stepwise regression of group (+1, 0, -1) scores on teaching performances for Grade 2 Reading. $R = .96$, $p < .0001$, and 16 teaching performance variables predict 92 percent of the variance in group scores. In other words, we can predict with great accuracy the pattern of fall-spring differences from information on these teaching variables. In the following sections descriptions of the analysis of "patterns of gains" refers to this analysis.

Residual scores, mean change scores, and pattern scores defined three different bases for analysis; multiple and stepwise regressions were performed with each type of score. Thus, six different basic types of analysis were performed to determine if and how teachers make an appreciable difference in how much pupils learned. All analyses were performed separately for each subject area -- reading and mathematics -- at each grade level. The separation of the analyses by grade level and subject area was done to determine if there were teaching performances which were effective in both grades and both subjects.

These three kinds of analyses were used to identify predictors of differences in pupil achievement which were described by three different kinds of scores. These analyses lead to a set of hypotheses about which teaching performances are likely to be associated with differences in pupil achievement.

Table 10

Stepwise Regression of Group Scores on Teaching Performance (Step 16)
Grade 2 Reading (N = 33)

```
THE DEPENDENT VARIABLE IS GROUPVAR,  THE MULTIPLE CORRELATION IS  0.9600,  THE STANDARD ERROR OF ESTIMATE=        0.0700

                                                  PROPORTION                                                   PROBABILITY
                              SUM OF SQUARES      OF SQUARES       N.D.F.     MEAN SQUARE       F RATIO        OF LARGER F
      TOTAL ABOUT ORIGIN           1.0000                             32.
   UNDER NULL HYPOTHESIS           1.0000            1.0000           32.
      DUE TO HYPOTHESIS            0.9215            0.9215           16.          0.0576        11.7411           0.0000
                  ERROR            0.0785            0.0785           16.          0.0049

                    STANDARD      REGRESSION      STANDARD     T STATISTICS    CONTRIBUTION     MEASURE OF
                   REG. WEIGHT      WEIGHTS     ERROR OF WT.  WITH   16.D.F.     TO R-SQ.      COLLINEARITY

INDEPENDENT VARIABLES

     APPLE--1       0.2217         0.2217         0.1025         2.1626        -0.0229         0.5331
     APPLE--7       0.0373         0.0373         0.0950         0.3929        -0.0008         0.4559
     RAW----1      -0.4756        -0.4756         0.1127        -4.2182        -0.0873         0.6140
     WD-----5       1.0353         1.0353         0.1190         8.7013        -0.3714         0.6535
     WD-----1      -0.9412        -0.9412         0.1199        -7.8508        -0.3023         0.6587
     WD-----3      -0.5693        -0.5693         0.0996        -5.7154        -0.1602         0.5057
     RAW----3       0.5455         0.5455         0.0898         6.0748        -0.1810         0.3917
     APPLE-12      -0.5559        -0.5559         0.1103        -5.0392        -0.1246         0.5969
     APPLE-10       0.4831         0.4831         0.1168         4.1347        -0.0839         0.6406
     APPLE--8      -0.3492        -0.3492         0.0966        -3.6134        -0.0640         0.4746
     APPLE-14       0.2785         0.2785         0.1017         2.7394        -0.0368         0.5256
     WD-----2      -0.4382        -0.4382         0.1041        -4.2079        -0.0869         0.5477
     APPLE--6      -0.2737        -0.2737         0.1101        -2.4852        -0.0303         0.5955
     APPLE--2      -0.3619        -0.3619         0.1206        -3.0013        -0.0442         0.6626
     WD-----4       0.2957         0.2957         0.1129         2.6180        -0.0336         0.6154
     APPLE--5       0.1790         0.1790         0.0889         2.0139        -0.0199         0.3790

OTHER VARIABLES                                           NOW    15.D.F.

     APPLE--4       0.1014         0.1014         0.1170         0.8667         0.0037         0.6359
     APPLE--3       0.0976         0.0976         0.1991         0.4900         0.0012         0.8701
     RAW----4      -0.0464        -0.0464         0.1175        -0.3949         0.0008         0.6252
     APPLE-11      -0.0420        -0.0420         0.1056        -0.3980         0.0008         0.5360
     RAW----2      -0.0807        -0.0807         0.1571        -0.5137         0.0014         0.7918
     APPLE-13       0.2567         0.2567         0.0973         2.6388         0.0249         0.6224
     APPLE--7       0.0096         0.0096         0.1122         0.0853         0.0000         0.5848
```

Such predictive hypotheses do not explain the relations among the variables. An attempt has been made in this study to develop the bases for an explanatory theory of teaching. We measured a number of domains of variables which were predicted to relate to teaching performances and pupil learning (see Figure 1). The structural model portrayed in Figure 1 suggested many questions whose answers would contribute to the development of a theory of effective teaching.

A fourth major kind of analysis was conducted to answer these major questions -- are teacher characteristics associated with differences in teaching actions and activities, and are characteristics of the teaching environment associated with differences in teaching performances? The procedure used to address these questions was to test a structural model of the relations among the domains of variables on which data were gathered.

A structural model is a description of the predicted causal relations among the variables. The formal model representing these predictions is a structural model, so called because the predicted relations may be tested by a structure of equations. Correlational techniques do not permit strict causal inference. But a structural model, which is tested statistically by using correlational methods, estimates the likelihood that a relation is causal. Estimating the likelihood of causal relations is a way of generating hypotheses about the relations among the variables -- a major purpose of this study. (In a study of this kind, it is not possible to make strict inferences about causality. A special set of experiments would have to be designed, in fact, to test causality.)

The necessary condition for predicting that two variables are causally related is that one of them occurs prior to the other. Teacher aptitudes, for example, are stable characteristics, that is, they are interpretations of the world that a teacher characteristically uses in a variety of situations. It may be logically assumed that these aptitudes existed prior to the specific behaviors observed; it is therefore reasonable to predict that aptitudes which already exist do determine individual teaching performances.

Although prior occurrence is a necessary condition for predicting a causal connection, previous research and theoretical analysis may also be used to support the prediction of a causal relation in a structural model. Previous research, for example, has shown that rewarding a behavior strengthens that behavior. Thus rewarding pupil behavior -- as a teaching performance -- should theoretically have a direct effect on what pupils do and learn.

A formal structural model is therefore a picture of hypothesized linear relations among the variables which are logically causal. The statistical procedure used estimates the constants for the equations representing these relations and tests how well the data fit the model. The technique used is path analysis. If the fit is good, the model can become a working theory which may be tested with other sets of data.

The use of such models is a way of organizing a large array of data. It is a way of making sense out of the data. Thus even if the data from the study do not fit the model, the attempt has been useful because one theory of the relations among the variables has been tested.

The structural model presented in Figure 1 postulates four different kinds of potential causal relations. The first of these is the influence of one person's behavior on another. The model predicts that teaching actions will directly affect pupil behavior: i.e., what a teacher does will directly

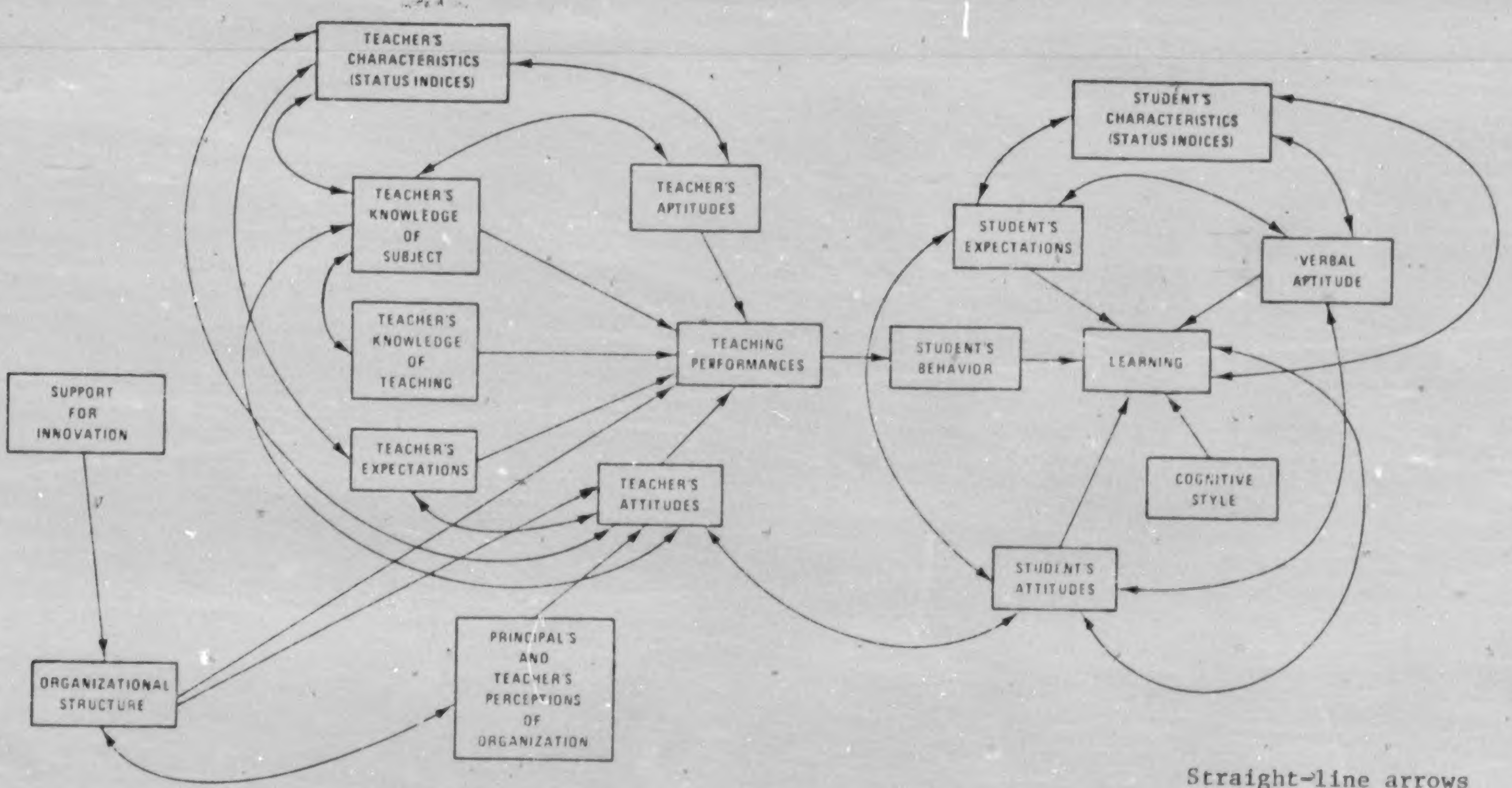

TEACHER'S CHARACTERISTICS (STATUS INDICES)
TEACHER'S KNOWLEDGE OF SUBJECT
TEACHER'S APTITUDES
TEACHER'S KNOWLEDGE OF TEACHING
TEACHING PERFORMANCES
SUPPORT FOR INNOVATION
TEACHER'S EXPECTATIONS
TEACHER'S ATTITUDES
ORGANIZATIONAL STRUCTURE
PRINCIPAL'S AND TEACHER'S PERCEPTIONS OF ORGANIZATION
STUDENT'S CHARACTERISTICS (STATUS INDICES)
STUDENT'S EXPECTATIONS
VERBAL APTITUDE
STUDENT'S BEHAVIOR
LEARNING
COGNITIVE STYLE
STUDENT'S ATTITUDES

Figure 1   A Structural Model of the Domain of Variables Influencing Teaching Performances and Children's Learning

Straight-line arrows represent predicted causal relations; curved arrows represent correlations but not causal relations.

influence what a pupil does in class (see boxes labelled "Teaching Performances"and"Student Behavior"). The teacher asks a question, a pupil answers it. The teacher organizes work, the pupils do it. A strong relation should be found between what the teacher is doing and what the pupils are doing. Since the teacher's actions usually, but not always, occur first, inferring causality from a substantial path coefficient is reasonable. Because different kinds of teacher behavior and several kinds of pupil behavior were being related in this study, variations in the size and significance of path coefficients should reveal those teaching performances which appear to have the most direct effect on pupil behavior.

The second kind of postulated causal relation is that between pupil behavior, which can be directly observed, and pupil learning, which can be measured only indirectly and incompletely (see boxes labelled "Student Behavior" and "Learning"). The greater a pupil's engagement with a learning activity, the more he or she is likely to learn. Since there were several measures of pupil behavior and of learning, it was possible to estimate the relative effects of the different behaviors on different kinds of learning.

But it is also likely that certain internal states or habits will affect learning directly, e.g., how much the pupil expects to learn or to be able to learn, the pupil's attitude toward what is to be learned, and his or her general aptitude for learning and learning style (see boxes surrounding "Learning" on the right hand side of Figure 1). The relative sizes of the path coefficients are estimates of these relative influences and, when compared to coefficients for the pupil behavior-learning path, will estimate the relative influence of behavior in the learning situation compared to the influence of these internal states and habits.

A third set of relations is of a similar character; that is, the set of relations between teachers' knowledge, attitudes, aptitudes and expectations, and actual teaching performance (see boxes on left side of Figure 1 in back of "Teaching Performances"). Here the expectation is that internal states and habits directly determine overt teaching behavior.

A fourth set of relations is that between classroom or institutional organizational structures and teaching performances (see boxes in lower left-hand corner of Figure 1). The antecedents in organizational structure include the degree of principal leadership, the products of administrative decisions such as curriculum choices, and the level of interactions among the staff. The events associated with different kinds of organizational structure are also expected to affect teachers' attitudes, and attitudes in turn are predicted to affect teaching performances.

Figure 1 presents a model of these relations. This model is a description of the nature of the relations, which are of two kinds, dependencies and associations. Dependencies are represented by straight lines, and associations by curved lines. Thus, the straight line from "Teacher's Aptitudes" to "Teaching Performances" is a pictorial way of saying that teaching performances depend on aptitudes; that is, how a teacher teaches depends on their perceptual-cognitive capacities. The curved line between "Teacher's Aptitudes" and "Teacher's Knowledge of Subject" is a pictorial way of saying that variables in these domains will be correlated.

Some may question why a dependent relation between knowledge and aptitude was not hypothesized. But to postulate a dependency relation, it is necessary to know that one of the variables occurs prior in time to the other, or there is a theoretical reason or there are experimental data to support such a stipulation. In the specific case of these two domains of variables we do not know which of the sets depends on the other; aptitudes may develop as a consequence of acquiring certain kinds of knowledge; or a developed aptitude may help a person to acquire a certain kind of knowledge. Since both kinds of dependencies are possible, and since we did not know which one was applicable to the specific aptitudes and knowledge measured in this study, we hypothesized an association rather than a dependency.

The purpose of using a structural model is to test a theory about dependencies. There are other theories than that represented in Figure 1; for example, the investigator who thinks that aptitudes are likely to have developed prior to the acquisition of formal knowledge will draw a straight line between the aptitude and knowledge boxes.

The point is that many different conceptions of dependency and association may be thought of and tested. The model that best fits the data is the best theory for that data. Because of restrictions of time we have tested only the model presented in Figure 1, a model which requires few assumptions about dependencies. We adhered rather strictly to two rules: 1) there must be reason to believe that one set of events preceded another in time; 2) there were theoretical reasons or empirical data to support a hypothesis that one set depended on another. When in doubt we opted to postulate an association rather than a dependency.

The reason for stipulating the dependencies and associations in advance is that there are statistical procedures which may be used to test the validity of the model. One such procedure is path analysis. The statistical methods used in a path analysis employ the zero-order and partial correlations among the variables in the various domains included in the model. The method yields path coefficients which are partial regression coefficients.

The methods of path analysis derive from the concept of partial correlation. The partial correlation method calculates the correlation of two variables independent of their correlations with other variables. In a path analysis this concept is applied to estimate the association between two variables hypothesized to be causally related. The path model reflects the theory of causal dependency; and the analytic procedures test this theory. The path coefficient is an estimate of the magnitude of the association between two variables hypothesized to be causally related. If the path coefficient is significant, we know the magnitude of the association and this estimate is independent of associations with other variables. Since we have reason to believe that the two variables are causally dependent, and since the path coefficient is significant, we infer a causal relation because the association between the variables cannot be attributed to their association with any other variable.

Note the arrow from "Teacher's Aptitudes" to "Teaching Performances." There are five aptitude scores in the domain, "Teacher's Aptitudes," and 22 scores in the domain, "Teaching Performances." There would be an arrow from each aptitude score to each teaching-performance score (a much too complicated diagram to read easily).

We found for Grade 2 Reading a path coefficient of .34 between "Flexibility" (an aptitude score) and "Teaching the Class-As-A-Whole" (a teaching-performance score). This path coefficient is an estimate of the strength of association between these two variables with the correlations between all the other combinations of aptitude and teaching-performance scores partialled out.

This coefficient takes into account, by "partialling out," the correlations of other aptitude scores with this teaching performance variable and with each other.

Suppose that we begin by looking at the zero-order correlations among the aptitude scores and the the performance variable, "Teaching the Class-As-A-Whole":

	TCH CL WH	VERB FL	MEM	REASN	FLEX
TCH CL WH	1.00				
VERB FL	.01	1.00			
MEM	.15	.45	1.00		
REASN	-.13	.48	.45	1.00	
FLEX	-.08	.68	.57	.63	1.00
COG STYLE	-.04	.54	.37	.43	.61

(TCH CL WH: Teaching the Class-As-A-Whole; VERB FL: Verbal Fluency; MEM: Memory; REASN: Reasoning; FLEX: Flexibility; COG STYLE: Cognitive Style)

Reading down the first column it can be seen that the largest zero-order correlation (r) is between "Memory" and "Teaching the Class-As-A-Whole." Reading down the other columns we note that the aptitude variables are substantially intercorrelated. If we made inferences only from these data, we would conclude that the aptitude variables were unrelated to this performance variable.

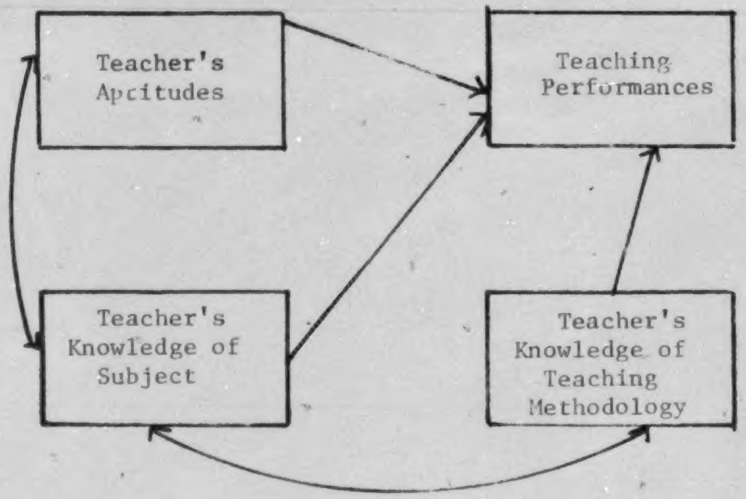
But, "Teaching the Class-As-A-Whole" is only one of 22 performance variables. It is correlated in varying amounts with other performance variables; for example, it is correlated with "Unsupervised Seatwork" (r=-.25) with "Teaching in Groups" (r=-.36), and with "Teacher Movement" (r=.33). Each of the other 21 performance variables is correlated in varying degrees with each of the aptitude scores.

The method of path analysis utilizes these correlations to sort out the magnitude of the association between two variables independently of their associations with other variables. Thus, when we obtain a path coefficient between "Flexibility" and "Teaching the Class-As-A-Whole" equal to .34, we know that this is the magnitude of the relation independent of the correlations of "Flexibility" with other aptitude variables, and of "Teaching the Class-As-A-Whole" with the other aptitude and performance variables.

But it is not the statistical technique of path analysis alone that makes it possible to make causal inferences. The model describes the hypothesized causal relations. It is a network of hypothesized dependencies and associations. We expect to find substantial and significant path coefficients for every relation portrayed by a straight line in the model. When we do not, a particular hypothesis has not been confirmed.

The process of reasoning to a causal inference may be illustrated by analyzing a part of the model in Figure 1. Note in Figure 1 that we have drawn a curved line between the domains of "Teacher's Aptitudes" and "Teacher's Knowledge of Subject," and we have also drawn a straight line between "Teacher's Knowledge of Subject" and "Teaching Performances." There is a third straight line between "Teacher's Knowledge of Methodology" and "Teaching Performances" and a curved line between the two "Knowledge" boxes.

These relations taken out of Figure 1 look like this:


Teacher's Aptitudes
Teaching Performances
Teacher's Knowledge of Subject
Teacher's Knowledge of Teaching Methodology

This model predicts that "Teacher's Aptitudes," "Teacher's Knowledge of Subject," and "Teacher's Knowledge of Teaching Methodology" have a direct causal influence on "Teaching Performances." The model also predicts an association (correlation) between aptitude and knowledge of subject (represented by the curved line), and between the two kinds of knowledge.

The theory portrayed here is that how a teacher perceives and conceives the phenomenal world (aptitudes), and what the teacher knows about the subject being taught and how to teach it determine how the teacher teaches. We will look at one specific kind of teaching action -- how frequently the teacher organizes the class as-a-whole for instruction -- as an illustration of this model.

Consider the following path coefficients:

PATH COEFFICIENTS FOR GRADE 2 READING

TEACHER'S APTITUDES	TEACHING THE CLASS-AS-A-WHOLE
	p
Verbal Fluency	n.s.
Memory	n.s.
Reasoning	n.s.
Flexibility	+ .34
Cognitive Style	n.s.
TEACHER'S KNOWLEDGE OF SUBJECT	
Knowledge of Methodology	n.s.
Knowledge of Subject	- .42

These data indicate that "Flexibility" *probably* has a direct causal influence on the teaching performance, "Teaching the Class-As-A-Whole," but the other aptitudes measured in this study do not have such an influence on this teaching performance. Similarly, there is no significant relation between the teacher's knowledge of methodology and this teaching performance; but there is a negative relation between knowledge of the subject being taught (reading) and this performance; that is, the more the teacher knows about reading, the less likely she or he is to teach the class-as-a-whole.

The conclusions from the above data are that the teacher who has higher scores on "Flexibility" is more likely to teach the class-as-a-whole, and the teacher who knows more about reading is less likely to teach in this way. Knowledge of teaching methodology is unrelated to the use of this instructional procedure.

It is not immediately obvious why these relations may be causal determinants of this teaching performance. The path analysis has enabled us to sort out the variables with the strongest relations to this teaching performance, and the inference that the relation is causal is based on what we know about the ordering of the variables in real time.

Since there is one other domain, "Teacher's Characteristics," predicted to be associated with aptitudes and knowledge, we may find relations which help us understand the dependencies among the variables. An analysis of the measures of the variables may give us clues about the processes involved. Or we may reorganize the model and test other patterns of relations. Thus the model and the path analysis technique are heuristic devices for developing a theory of effective teaching.

As the results of the path analysis are studied, it is important to keep in mind the results of the regression analyses. We know, for example, that teaching the class-as-a-whole is unrelated to differences in pupil achievement. So the relations we have been discussing are directed to finding who is more likely to use such a practice and why.

The logic of the data analyses is represented in the preceding pages. We first try to find among teaching performances the best predictors of differences in pupil achievement. We then try to find by the analysis of a structural model the probable "causes" of differences among the teachers in the way they teach.

The results of this study described in the following pages are of two kinds. First, we describe which teaching acts and activities were associated with greater changes in children's learning. These associations form the basis for hypotheses that may be tested in subsequent phases of the study. Second, we present the results of the testing of a structural model. This analysis identifies the variables likely to influence teaching performances. This information may be used to design more sophisticated models for future research. The testing of a model of this kind contributes to a theory of teaching by suggesting what variables should be included in such theories and by demonstrating the relations among them.

THE RESULTS OF THE STUDY

Several general conclusions of pertinence to questions of perennial interest in research on teaching were derived from the analyses. This research, for example, has long sought "basic" teaching skills that significantly influence learning. Basic skills have been thought of as techniques of teaching that would promote learning in all subjects. "What makes an effective teacher?" has frequently been asked as if there were a single answer. But one important result of this study is that no such single skill or teaching performance was found to be equally or comparably effective in both grades or in both subjects.

This result, if replicated or if found in comparable studies, is important. Its practical implications are that the goals of training teachers for the primary or the intermediate grades and the criteria for evaluating their competence will necessarily be different.

It was also found that relatively few teaching-performance variables when considered alone related significantly to either type of pupil change score. This result suggests that *there are no single teaching-performance* variables correlating so significantly with children's learning that they should be considered critical for effective teaching. This result should influence the direction of future research. In the past most research on teaching was the study of the zero-order correlations of a teaching act or event or activity with a measure of learning. It is well-known that this mode of analysis has yielded meager results. This study suggests that it is different patterns and structures of teaching acts that influence changes in learning rather than single, omni-effective teaching performances.

A third general conclusion is of particular importance. Teacher effects proved significant in a substantial proportion of the analyses in this study. Although it is true that pupil scores in the fall predicted 70 to 80 percent of the variance in spring scores, significant correlations were obtained when spring scores, with fall scores partialled out, were regressed on teaching performances (residual scores regressed on performance scores) and when mean difference scores were regressed on teaching performance scores.

In these analyses, teaching performance variables accounted for 11 to 64 percent of the variance in the change scores; correlations for mean change scores ranged from .33 to .80. When residual scores were used in these analyses, correlations ranged from .56 to .72; the performance variables accounted for 31 to 52 percent of the variance remaining after fall scores were partialled out.

These results, particularly those derived from the analyses of residual scores, are significant for developing a theory of effective teaching. The major aim in building this theory is to describe how specific teaching actions or styles affect different kinds of learning. This aim cannot be achieved if we cannot link the acts of each teacher with the classroom behavior and learning of his or her students. Also we must identify how much of a pupil's learning is attributable to a specific teacher.

The design used in this study, for example, used pretest measures which located pupils on different achievement and aptitude scales, assessed their expectations and attitudes, and collected information on their previous educational history and socioeconomic background. This type of design permits us to draw conclusions about the influence of these teachers on the children they taught during the year of this study.

The critical empirical question is, of course, "Is there a relation between teaching performance scores and learning scores?" The magnitude and significance of the multiple R's indicates that there are sets of teaching performance variables which are significant predictors of differences in pupils' achievement scores. But how sure can we be that there might be a causal connection between teaching acts and learning, and what is the magnitude of the effect?

As is well known, a causal inference cannot be made directly from a significant correlation between two variables. The complementary principle is, however, frequently ignored, namely, that when a significant correlation is found a causal relation exists somewhere -- perhaps between the two variables or between the two variables and a third, or any of a number of other possibilities. Although these logical possibilities can be thought of, the design of a study allows us to eliminate some of them as implausible, and to infer that others are more plausible.

In this study, for example, we know how much of the spring reading and mathematics score variance is predicted by the fall score variance -- 64 to 81 percent. These fall scores are correlated with a variety of socioeconomic status indicators, and they are measures of the pupils' previous learning. Therefore, we know how much of the differences in pupils' achievement is predicted by these factors and may reasonably be attributed to them. Few will argue that the pupils' previous learning and the conditions that stimulate interest and support learning are not causal factors influencing learning. If, therefore, we find, as we have, that teaching variables predict differences, it is plausible that differences in teaching acts are causally linked to differences in achievement.

But there is at most one-third of the remaining variance to predict, and consequently that may be attributed to the actions of the teacher. Table 11 summarizes the R and R^2 values found in the various analyses and the percentage of the variance predicted by the teaching performance variables. Note particularly the percentage of variance predicted in the analyses in which residual scores were used. (Refer to the columns under "Discriminant Group." Estimates for the classes in this set are more likely to be reliable.)

The pretest scores predict on the average about 70 percent of the variance in spring scores. There is only 30 percent of the variance remaining to predict. The teaching performance variables predict a third to half of this remaining variance. Some of this remaining variance is attributable to error of measurement.

The logic of the reasoning is as follows: (1) assume that the reliability of the posttest is .90; then approximately 80 percent of the variance can be

Table 11

Summary of R, R^2 and Percent Variance Accounted for in Stepwise Regressions of Residual, Mean Gain and Pattern Scores on Teaching Performance

	TOTAL GROUP			DISCRIMINANT GROUP		
Reading						
Grade 2	(N = 40)			(N = 33)		
Residual Scores	.59	.35	35	.68	.46	46
Mean Gain Scores	N.S.	---	--	.77	.59	59
Pattern Scores	---	---	--	.96	.92	92
Grade 5	(N = 53)			(N = 41)		
Residual Scores	.56	.31	31	.59	.35	35
Mean Gain Scores	.33	.11	11	.39	.15	15
Pattern Scores	---	---	--	.64	.41	41
Mathematics						
Grade 2	(N = 41)			(N = 37)		
Residual Scores	.67	.46	46	.67	.45	45
Mean Gain Scores	.80	.64	64	.58	.34	34
Pattern Scores	---	---	--	.61	.37	37
Grade 5	(N = 54)			(N = 45)		
Residual Scores	.72	.51	51	.70	.49	49
Mean Gain Scores	.79	.62	62	.65	.42	42
Pattern Scores	---	---	--	.60	.36	36

predicted by other factors; (2) the pretest score predicts about 70 percent of this 80 percent; (3) teaching performance variables predict a third to a half of the remaining variance to be predicted.

What is more interesting than this general case is the analysis of differences in achievement by class. These differences are estimated by what we have called "pattern scores." Refer again to Table 11 in the columns under "Discriminant Group" and in the rows across from "Pattern Scores." Teaching performances at Grade 2 and in Reading predict almost entirely (92 percent) the differences in patterns of achievement. In other subjects and grades the percentage of the variance predicted is 35 to 40 percent.

These data give us an idea of the magnitude of teacher effects in terms of a prediction model. Such data are useful for developing a theory of effective teaching. But how much of a difference in learning do these teachers make?

In the following discussion reference to Figure 2 will be useful to the reader. The teacher in the upper left-hand corner, T22, represents a class in which all children in the class made gains over the course of the year. The lowest fall score in this teacher's class was 78; the lowest spring score was 146. By using the regression line for this teacher's class and by checking a manual for a standardized test, we estimated that, had we been using norm-referenced tests, children at the lower end of the scale in the fall would have gained no less than two grade levels and as much as three grade levels.

By way of contrast, consider the class represented by T11 in the lower right-hand corner. In this class pupils at the higher end of this classes' fall scores made greater gains than those at the lower ends. The ranges for fall and spring scores are as follows:

R_f: 99-206 R_s: 68-215

The differences in pupils' achievement in T22's and T11's classes cannot be explained by differences in the socioeconomic status or aptitudes of the pupils.

The results with the three different kinds of scores indicated that teaching performances make a substantial contribution to what children learn. These claims should not, however, be interpreted to mean that the correlates of initial position (e.g., the socioeconomic status of the learner) are irrelevant. In the analysis of the data the factor most strongly related to a pupil's final score was initial scores, and these initial scores correlated positively with socioeconomic status. In general, pupils who read better than their classmates at the beginning of the year were still the better readers at the end of the year. But in this study, the second most important factor predicting change was what teachers did. Initial position plus teachers' actions accounted for almost all the variance in final scores.

Practically these findings mean that a pupil may make more progress during the school year than would be expected from the level of his or her skills in September -- provided that the teaching received is like what was found to be effective in the Phase II study. Such changes occurred with pupils whose initial scores were low compared to his or her classmates', as well as with pupils who were above average or better in reading or mathematics.

A fourth major conclusion was that differences in patterns of teaching performances accounted for differences in pupil learning. In this study, 15 of 16 stepwise regressions at both grade levels in reading and mathematics yielded significant multiple correlations. The stepwise regression takes into account the relative contribution of all teaching performance variables with respect to each other. When the set of positive and negative predictors found in these analyses was studied, they described different and contrasting styles of teaching. Therefore, it is reasonable to conclude that, in a study of teaching effects, one should be seeking patterns of performance that affect children's learning.

The data suggest two hypotheses: first, a pattern of teaching performances is more likely to be related to learning than is a single performance; second, effective teaching patterns differ by subject matter and by grade level.

The results also lend support to hypotheses about which patterns of practices may account for differences in relative effectiveness. The teachers in this study, for example, used various ways of organizing pupils for instruction. The three most typical patterns were working with an individual pupil, teaching a group of pupils, and teaching the class-as-a-whole. Some teachers consistently used one of these; others used a mixture. The data suggest that no one method of classroom organization was consistently effective at both grade levels in both reading and mathematics.

But there appeared to be a structure in effective teaching practices. If the pattern of classroom organization and teaching performances provided more direct instruction to pupils by the teacher, the learning of the pupils was improved. The more effective teaching performances provided more direct instruction to pupils by the teacher. The significant teaching performance variables for each grade level and subject will be described in the sections that follow. We use the stepwise regression analyses in these descriptions because they typically provided the best prediction of differences in achievement. We also provide the results from the analyses of pattern scores.

Grade-Two Reading

A stepwise regression in which spring scores were regressed on teaching performance variables with fall scores partialled out (residual scores) yielded a multiple correlation of .59 ($p \leq .01$). Performance variables accounted for about 35 percent of the variance in spring scores when fall scores were partialled out.

One positive predictor was obtained in this analysis, a variable that is a measure of the degree to which a teacher used different kinds of teaching materials. Two negative predictors appeared, a measure of the complexity of the organizational structure of the classroom, and a measure of the extent to which the teacher attempted to achieve many different instructional goals during the instructional period.

Recall that two patterns were found in the mean difference scores: one of large mean differences spread more or less evenly across the class, and the second of small mean differences, the greater changes being made by pupils with higher fall scores. In the second pattern, fall scores were the major predictor of spring scores. In the first, however, fall scores were a weaker predictor of spring pupil performance. Teaching performances accounted for these differences in the pattern of mean change. In a stepwise regression of scores associated with these patterns on teaching performances, R = .96 ($p \leq .001$), and $R^2 = .92$; that is, differences in teaching performances accounted for 92 percent of the variance in the pattern of mean change scores. The differences in the patterns were accounted for almost entirely by teaching performance and pupil behavior variables.

The teaching pattern reflected in the positive predictors of pupil change (the first pattern) was one in which children were placed in independent seatwork, but the teacher interacted with them by initiating discussion, explanations and questions, kept them on-task by redirecting them to the task, and used a greater variety of instructional materials.

The negative predictors (pattern two) suggested the following picture: the children were taught in groups; there was a higher level of pupil inattentiveness; the teaching activity involved more organizing of instruction than imparting direct instruction; the teachers responded to pupil questions or to pupil behavior rather than initiating the interaction themselves; and the teachers taught several kinds of content, each requiring specific strategies, materials, and activities--a difficult arrangement to manage.

These positive and negative predictors suggest contrasting teaching styles. Teachers whose scores were higher on the positive predictors were more frequently observed in those performances; whereas teachers whose scores on the negative predictors were higher were more frequently observed in those performances. This analysis suggests that the patterns of teaching performances may explain differences in the patterns of pupil changes.

Second-grade reading is devoted typically to acquiring basic decoding skills and the rudiments of reading comprehension. Pupils frequently differ considerably in how much decoding skill they have acquired. Consequently, teachers organize instruction to accommodate the differences. Much independent seatwork and teaching of small groups were observed at this grade.

The data suggested that this pattern of instruction was most effective when it contained the following elements: (1) each pupil's opportunity for prompt access to the teacher for instruction or for correction of work was increased as much as possible; and (2) when the teacher worked with an individual pupil or with a small group, the other pupils persisted in their work on reading tasks. Unmonitored seatwork and arrangements that kept many pupils waiting for instruction were ineffective. The arrangements thus resemble the tutorial model, except that all pupils in the classroom must be instructed.

These results seem to be consistent with what is known about how basic skills are learned. The skills must be demonstrated or practiced and the pupils must be told when they are making errors. The teacher, for example, may model a correct sound, or have the pupils practice it, correcting them as they produce the sound. The more this kind of direct instruction is provided for each pupil, the more each learns.

Grade-Five Reading

When residual scores were used in a stepwise regression analysis, R = .56 ($p \leq .02$) and $R^2 = .31$; the latter statistic indicates that performance variables accounted for 21 percent of the variance in spring scores when fall scores were partialled out. Note again that this percentage is a substantial contribution in the variance analysis.

In this analysis, two positive predictors were obtained: (1) a measure of the extent to which the teacher discussed, explained, questioned, and answered in a sustained fashion; and (2) a measure of direct individual instruction. Negative predictors of pupil learning were the use of a diversity of materials and pupil inattentiveness.

Reading comprehension is usually stressed during fifth-grade reading instruction, and pupils are asked to read increasingly complex materials. For full comprehension, this reading material requires complex cognitive processes, such as analysis and evaluation. The teacher's task in reading instruction is, therefore, to stimulate higher levels of thinking. Teaching practices which provided for sustained interactions with pupils were more effective for teaching fifth-grade reading probably because they stimulated the thinking needed for improved comprehension. Such practices included conducting discussions, giving explanations, and sustained questioning and answering.

The contrast with effective practice in second-grade reading instruction is marked. At the second-grade level, the pupil is learning discrete responses which must be practiced or is using simple cognitive skills such as identifying the principal features of a story. Frequent encounters between the teacher and pupil, in which the teacher demonstrated the skill or monitored practice, were an effective method of direct instruction. But in the fifth grade the teacher must think with the pupil about what he or she has read, what it means, and how it relates to other ideas.

It is interesting that in grade 5 the use of a variety of teaching materials was a negative predictor. The instruction needed to improve reading comprehension at this level requires sustained interaction with pupils. But the teacher who uses a diversity of materials may in effect be reducing the amount of time for sustained work with different children on any given set of materials. If this speculation has some validity, a critical factor may be pupil and teacher interaction with the materials rather than the diversity of the materials themselves. Presumably pupils are learning complex skills. Although one suspects that the material has to be appropriate to the interest and level of skills of the pupil, too great a diversity of materials may interfere with the basic instructional activity required to achieve the instructional goals.

The classes were sorted by patterns of pupil gains. Stepwise regression of scores associated with patterns yielded an $R = .64$ ($p \leq .001$) and $R^2 = .41$; therefore, teaching peformances accounted for 41 percent of the variance in pattern scores.

In this analysis, the amount of time spent teaching and in direct individual instruction were positive predictors. Time spent by the teacher in organizing instruction, an activity that is the opposite of direct instruction, and a complex organizational structure were negative predictors.

The results for grade 5 reading were remarkably consistent and their implications clear. If one were to generate an algorithm for teaching reading at this grade, one should assign most if not all of the instructional time to individual reading. This reading should be discussed with each student by the teacher, who will use explanations and questions to stimulate thinking. If groups are organized, they should be used for discussions of reading materials. But a group should not be used as an alternative to direct individual instruction.

Grade-Two Mathematics

The stepwise regression of spring scores on teaching performance variables with fall scores partialled out (residual scores) yielded an $R = .67$ ($p \leq .01$), and $R^2 = .45$ when six variables were used. These six variables accounted for 45 percent of the variance in spring scores when fall scores were partialled out; four had significant regression weights. In this analysis, the only positive predictor was the variety and number of mathematical skills taught.

The negative predictors were an interesting combination of variables: teaching the class-as-a-whole, using a complex organizational structure, and unsustained reactions to pupil activities. These three variables describe teaching in which (1) the teacher uses a complex organizational arrangement for teaching, such as alternating individual and group instruction but teaching the class-as-a-whole more frequently, and (2) the teacher uses a pattern of questioning that is a series of responses to children's questions or work rather than instruction through initiating questions, offering explanations, or discussing work.

This type of analysis suggests that the mathematical knowledge a child brings to the class (fall scores) plus the amount of content that he or she is taught are the best positive predictors of where the child will be performing by the end of the year. The negative predictors probably describe less effective ways to teach second-grade mathematics.

It is useful to recall that teaching mathematics in the second grade commonly proceeds somewhat as follows: the teacher stands before the class and explains a problem or a concept or a way of performing a mathematical operation. The children will have either worked on the problem or operation prior to this moment of instruction or are about to. As the teacher explains, he or she asks questions of different students in the class about the material being taught. This method of instruction, albeit typical, was not effective.

Mathematics instruction at the second grade is fairly concrete. The students are learning the basic concepts of common arithmetic and are performing a series of operations which enable them to combine various kinds of numbers. The process of combining numbers requires a series of discrete steps, the retention of "basic facts" such as the times-tables, and careful checking of the operations performed. The content of second-grade mathematics is more likely to be learned if the student repeats these operations in as many different ways as possible and receives corrective feedback on the operations performed either from the teacher or from the instructional materials.

The performances of the operations are also ways of understanding the underlying concepts, such as the union of sets, the basic concept underlying the processes of addition. Understanding the concept is related to the ability to perform the operations; for example, a child may be able to combine groups of familiar things such as blocks or other small objects; but if the child is to understand mathematics, she or he must move from the concrete phenomenal world to the abstract world of quantities. The child must learn how to combine these abstract quantities.

In order to acquire both understanding of and skill in these operations, the child must practice specific combinations in a variety of forms. It would appear then that the more direct experience the child has in these operations, the more likely the learning. It is also reasonable to assume that if the teacher carefully monitors the processes of this experience, the corrective feedback to pupils will facilitate their acquiring skill in combining quantities.

A reasonable inference is that children who are taught more content learn more mathematics over the year, and the amount learned is correlated with what they knew at the beginning of the year. But differences in how much is learned by whom depends on the instructional methods the teacher uses. Direct individual instruction produces greater differential change for more children. It should not be assumed, however, that greater learning for more children requires that less content be covered. Obviously the problem is to develop a methodology that increases both the amount of content covered and the amount of direct individual instruction.

The data from the two kinds of analysis suggest that any performance will be negatively related to learning if it either interferes with a system in which the child can work on operations at the level required to learn them or is a substitute for a system that encourages the performance of these operations. Teaching the class-as-a-whole is a system which, while it is being used, consumes time that the child might more profitably be spending in direct performance of mathematical operations and understanding and applying concepts. Similarly, an interactive mode in which the teacher asks a child a question and then moves to another child for a correct answer neglects to keep a particular child's attention focused directly on the processes or the operations until mastery of a combination or concept occurs.

Grade-Five Mathematics

The stepwise regression of residual scores on teaching performances yielded a multiple $R = .72$ and $R^2 = .51$. Teaching-performance variables accounted for 51 percent of the variance in spring scores when fall scores were partialled out. All of these variables, except one, had significant regression weights, and therefore this analysis yielded a prediction equation from which conclusions about effective teaching performance may be drawn.

When the classes were sorted by *patterns of pupil changes*, the stepwise regression of scores associated with the patterns yielded an $R = .61$ ($p \leq .01$) and $R^2 = .37$; therefore teaching performances accounted for 37 percent of the variance in *pattern scores*.

In this analysis, directed pupil seatwork, the quality of teacher planning for instruction, and teacher mobility in the classroom were positive predictors. The picture is one that suggests the benefits of more individual teaching of pupils. The data analysis also showed, in contrast to the results of stepwise regression of residual scores, that teaching more skills was a negative predictor.

There are two possible explanations for this difference in results between the analyses. The first is technical. In stepwise regressions a variable will take on a negative sign if it is positively correlated with variables which are positive predictors and are entered first into the regression analysis. The variable, "teaching more content," was significantly and positively correlated with the variable which had the largest regression weight and which had been entered first in the analysis.

The amount of content covered was a positive and significant predictor when initial scores were partialled out; that is, it was an effective variable when taken in combination with what a pupil already knew. But it was ineffective when *differences in the amount of the change* were considered (mean change scores); that is, larger average changes were associated with less content covered. These differences may mean that adding more new content is more effective with some pupils than others. But the average change in learning is not so affected because some pupils may profit more from what is added than others may. These data suggest that all pupils who are taught more content may profit from it, but how much they gain does not depend on the sheer addition of more content.

82 A

The positive predictors were these: varied organizational arrangements, amount of time spent by the teacher teaching a group, amount of time spent by the teacher teaching the class-as-a-whole, and a teaching performance that is a form of social control directed to keeping the children on-task.

Two modes of instruction appeared effective with the fifth grade that were not effective with the second grade. Apparently in teaching fifth-grade mathematics the teacher may effectively teach either the class-as-a-whole or groups. Several reasons may explain these differences by grade level. Presumably by the fifth grade pupils have mastered or been familiarized with the basic mathematical operations. The emphasis in fifth-grade mathematics is on concepts and problems, and the operations being learned are variants on operations learned in the earlier grades. Pupils may not need as much direct instruction on concrete specific operations. Because fifth-grade children have also had more experience with instruction, it may be that they can learn more easily in larger units than second graders can.

Using a variety of materials was again a negative predictor as was pupil inattentiveness, pupil attentiveness, and pupil productivity. The negative prediction associated with using a variety of materials may be a proxy for unstructured content. It may be that when teachers use a variety of materials the child literally jumps from one kind of task to another without seeing the relevance of and connection between the tasks.

Recall again that classes received scores based on the pattern of pupil changes. When teaching performance scores were regressed on these pattern scores, $R = .60$ ($p \leq .02$) and $R^2 = .36$; thus teaching performances predicted 36 percent of the variance in scores which described the pattern of mean changes.

In this analysis, the organizational complexity of the classroom was the only positive predictor. A teacher scoring high on this variable varied the context, teaching the class in groups or the whole class simultaneously and also providing some individual instruction. Teachers who varied instructional contexts produced a pattern of mean changes in which the gains were spread across most or all of the students.

The absence of other positive predictors should be noted. Simply varying the context of instruction is not likely to be the major determinant of how much children learn. Grouping is likely to be effective when the teacher creates an analogue of direct individual instruction, when pupil attention is maintained, and when the interaction is sustained through discussion or explanation or systematic questioning. In the analyses for fifth-grade mathematics, such variables were found to be positive predictors. It was the variable, variation of instructional contexts, that predicted a spread of gains across students, but the teachers must have also provided the other conditions of effective instruction.

This explanation is consistent with the picture that appears in the negative predictors that were found significant in this analysis. Assuming a variety of teaching roles, organizing for instruction, and using a diversity of materials were all negative predictors. A teacher who had high scores on these three variables created conditions so that the children who learned the most were those who originally knew the most.

Of particular interest in the fifth-grade analyses was the appearance of the organization-of-instruction variables--teaching in a group and teaching the class-as-a-whole--as positive predictors of learning changes.

It may be that it is easier to control the task-orientation of older children by using these procedures. Or instructional materials for the fifth grade may not be sufficiently self-instructional, so that a teacher who uses independent work must provide additional components of instruction, such as corrective feedback. If such is the case, then teachers will find it difficult to provide these for all students; grouping them or teaching the class-as-a-whole helps spread instruction to more pupils.

Another possibility is that the concepts taught in the fifth grade are better learned if the teachers spend more time explaining and discussing them. If so, then grouping the children is an efficient instructional procedure.

Suggested Hypotheses

If one were to put together the results obtained in the various analyses, what would be some "rules of thumb" about what to do and what not to do in teaching reading and mathematics? Below are some suggestions that should be treated as hypotheses:

1. Maintaining the on-task behavior of pupils is important, particularly for the fifth grade. Some forms of organizing instruction, such as teaching in a group, may lend themselves to increasing attentiveness because the teacher can exercise more direct control over pupil behavior. But with the second grade, although teaching in the group may lead to greater attentiveness, it is an ineffective teaching procedure. The problem for the second-grade teacher is to maintain the on-task behavior of all pupils while also providing a maximum amount of direct individual instruction.

2. Using a variety of materials is an effective strategy only in teaching second-grade reading probably because of the nature of what is being learned.

The use of a variety of materials was not effective for teaching fifth grade. There the materials may be inadequate substitutes for direct instruction. They may increase the complexity of the learning task or they may interfere with the management of instruction.

3. The variables that appear to be critical for teaching second-grade reading are (1) providing as much direct individual instruction as possible; and (2) concomitantly permitting the child to work with a variety of materials.

4. Direct individual instruction is also a critical variable for teaching second-grade mathematics. Planned sequences of teaching activities and attention to the problems of practice and transfer of learning may be the underlying critical variables.

5. Sustained interaction with children on reading materials and direct individual instruction are probably critical in teaching fifth-grade reading.

6. Teaching in a group or teaching the class-as-a-whole are effective ways to organize instruction in fifth-grade mathematics. But such modes of organization assume that the teacher maintains control of pupil behavior in these contexts.

7. Indirect forms of social control such as redirecting a child's attention appear to be the most effective ways of maintaining productivity and on-task behavior.

8. Interactive teaching performances which are brief or reactive in character are usually ineffective.

9. The expenditure of a good deal of time organizing for instruction is not an effective teaching strategy.

There is a theme common in these "rules of thumb": if an organizational context or a teaching performance provides for direct instruction, learning is improved. Thus, research should study teaching performances as mediators of variables that characterize a model of direct instruction. The critical problem, probably, is not whether a teacher teaches in a group or not. It is to what degree such teaching practices allow for direct instruction. When a particular teacher performance is found to be nonproductive, as was the case with teaching the class-as-a-whole in grade-two mathematics, it should be examined to find out what is missing from the model of direct instruction.

Direct instruction includes: (1) a component which explains what is to be learned or models it or elicits its elements by questioning; (2) a component which provides the appropriate conditions for attempting what is to be learned; and (3) a component which provides feedback on how well the child is learning what is to be learned. These components are most effectively provided in direct individual interaction with the child. If direct interaction is not possible, substitutes or alternatives should provide for these components in some other way. Independent seatwork is ineffective when the teacher does not provide sufficient monitoring or reduces the amount of explanation or elicitation of relevant responses.

It is important to emphasize that the data of the Phase II study describe characteristic styles or predominant modes of teaching. The results of this study suggest which styles or components of styles are likely to produce learning. It is the combinations of teaching performances (styles) rather than any single component performance that make for effectiveness of instruction.

SUMMARIES OF SPECIAL STUDIES WITHIN THE TOTAL STUDY

Several special studies were conducted and reported for Phase II. The studies are of two general types: those in which segments of the information included in the major analyses of the study are examined more intensively, and those in which ancillary research questions were explored. Included were extended reports on the two systems used for live observations of teaching performance in the project classrooms; a description of teaching performance based on teachers' reports of their own instructional activities; a study of teacher aptitudes, knowledge and attitudes as predictors of teaching performance; and, a discussion of the relationships among teacher expectations and pupil performances. The ancillary research questions were addressed in reports of the evaluation of teaching behavior through observation of videotapes, the gathering and analysis of historical test data about previous pupil performance in the classrooms of the Phase II teachers and the development and pilot testing of a videotaped test of teachers' skills in diagnosing reading problems.

INTENSIVE STUDIES OF THE MAJOR PHASE II SOURCES OF DATA

APPLE Observational Variables as Measures of Teacher Performance and Pupil Behavior

The APPLE[12] Observational System was one of two observational systems used in the study. The APPLE System involved a written record of the behaviors of teachers and pupils. "Scores" in the form of prorated frequencies from the observations were used as measures of teaching performance and pupil behavior in the study.

---

[12] APPLE Observation Variables and Their Relationship to Reading and Mathematics Achievement, Nadine M. Lambert and Carolyn S. Hartsough, Volume III, Chapter 1, Phase II Final Report.

The APPLE Observation System is a pupil-based system for recording behavior events in the classroom. The basic unit of observation is a pupil event -- what the child is doing when observed. Records of pupil behavior carry with them mandatory information describing pupil activities, the organizational structure of the classroom (instructional context), the teacher's instructional activities and the teacher's responses to the pupil behavior.[13] Observers make English language descriptions of these events with mandatory information which are subsequently encoded by reference to APPLE Lexicons.[14] The APPLE Lexicon was derived from actual observations of pupil and teacher behavior and therefore represents real, rather than hypothesized, classroom events. Examples from the Lexicon for each category are provided in Table 12.

APPLE differs from most other observation systems in that it is a low-inference naturalistic observation process which retains the pupil behavior annotations. Additionally, it is a system which readily yields categorical observational data. When treated as a category system, APPLE observation records are accumulated to measure the variety of naturally occurring pupil behaviors, pupil activities, teacher activities, instructional contexts and teacher responses using either an individual pupil or the entire class as the unit of analysis.

---

[13] APPLE: Manual for APPLE Observers. Lambert, N.M., Hartsough, C. S., Caffrey, C. & Urbanski, C. Berkeley, California: University of California, 1976.

[14] APPLE: Lexicon for APPLE Observations. Lambert, N. M., Hartsough, C. S., Caffrey, C. & Urbanski, C. Berkeley, California: University of California, 1976.

Table 12

Examples from APPLE Lexicon for Each Category

Instructional Context

Adult-Group: The class is divided into groups and an adult other than teacher directs activity of the group.
Independent-Class: The entire class is working or playing independently; teacher usually present but not directing.
Independent-Group: The group is working or playing independently of the teacher's direction.
Independent-Individual: The student is working or playing independently of the teacher's direction and of the rest of the class.
Teacher-Class: Teacher is directing the activity and the entire class is participating.
Teacher-Group: The class is divided into groups and the teacher is directing the activity of one group.

Teacher Activities

Academic Organizing: Giving directions, explaining and doing other activities which direct the students into activities.
Answering: Teacher answers question(s). A single answer to a single question or answers to questions put by several students over a period of time.
Asking: Teacher asks questions of a single student, a group, or the class.
At Desk: Teacher works at desk; activity is unspecified or unknown or when she is available to help students at her desk.
At Board: Teacher is working at blackboard, either alone or with student(s).
Checking: Teacher is checking work to see if material is understood, to see if instructions are being followed, to see how well work is being done, or if work is correct.
Circulating: Teacher is circulating around room, interacting with students, asking and answering questions, giving help.
Discussion: Teacher is leading a discussion. Emphasis on student response and comprehension, not simply on facts.
Explaining: Teacher is explaining something, either in response to a question or as part of giving instruction.
Helping: Teacher is helping a pupil or pupils in unspecified manner.
Instruction Giving: Teacher is giving instructions either about an assigned activity or about classroom organization.
Listening: Teacher listens to a student, either during oral reading or while student speaks to her or asks a question.
Question and Answer: Teacher is asking a series of questions, which students answer. Distinguished from ASKING by serial nature of questions.
Supervising: Teacher watches over class or a group. May occasionally ask or answer questions.
Working With: General unspecified instructional activity of working with an individual, a group or the class.

Teacher Responses

Teacher Help: Teacher gives some specific instructional assistance to an individual student.
Moves On: Teacher does not comment on a pupil response; instead, she calls on another pupil or moves on to another part of the lesson.
Positive Feedback: Use of mild positive verbal reinforcement.
None: No response to pupil performance or behavior. No distinction made between whether teacher is aware of this performance or behavior or not (unless is clearly a situation of IGNORING).
Redirecting: A verbal or non-verbal action by the teacher to redirect a pupil's behavior. Done without apparent anger.
Recognizing: Teacher's neutral nonevaluative response to an academic performance or a nonacademic behavior, e.g., "Notices," "acknowledges."

Pupil Events

+Attentiveness: Act of paying attention to the teacher teaching a subject, as shown through physical orientation, gestures, verbal response.
-Talking: Inappropriate or forbidden speaking.
Inattention: Short-term distraction from activity being directed by teacher.
+Oral Reading: Positive quality of student's reading out loud; fluency, accuracy, expressiveness, etc.
+Phonic Skills: Positive evidence of skill in sounding out new words; knowing the sound connected with letters.
+Reading Comprehension: Pupil's correct understanding of what he has read.
+Academic Performance: Level of performance in school tasks; response to assignments or lessons. Correct.
+Direction Following: Behavior related to following directions given by the teacher. Cooperation with teacher directives, whether academic or behavioral.

Copyright © 1971, 1976 by Nadine M. Lambert

The computer software which was developed for the APPLE system[15] provides combinations of the observation categories to yield measures of dyadic classroom variables such as Instructional Context-Teacher Activity, Pupil Event-Teacher Response, or Teacher Activity-Teacher Response. In addition to providing counts of frequencies of events, the APPLE system also incorporates a time-recording feature to measure the duration of classroom behavioral events.

APPLE in the Beginning Teacher Evaluation Study. APPLE observers recorded teacher and pupil behavior during reading and mathematics instruction for a total of 200 days for the 95 teachers in the Beginning Teacher Evaluation Study (McDonald & Elias, 1976). During these observation sessions, the observers recorded the behavior of eight target pupils in each class. These pupils were selected to represent varying achievement levels, different teacher expectancies for achievement, both sexes, and the ethnic distribution of the class.

These observations were encoded using the APPLE Lexicon and then were analyzed to determine the most frequently appearing instructional contexts, teaching activities, pupil behaviors, and teacher responses. This analysis produced 50 variables (six instructional context, 15 teaching activities, 17 pupil behavior, and 12 teacher response variables). The prorated frequency of each of these 50 variables was then computed within subject matter for each teacher.

---

[15] APPLE: An annotated pupil information system. Lambert, N. M., Hartsough, C. S., Converse, F. L. & Converse, E. K. Educational Technology, November 1971, 32-36.

Since we wished to obtain measures which would describe the classrooms in our sample fairly, a method of treating the raw observation frequencies had to be found which would equalize them for differential observation rates due to variability in the assigned number of observations in the classrooms or to differences in the extent of classroom structure. Our solution was to prorate the total frequency of each observation variable with respect to the total number of observations recorded overall (across all days of observation) for each teacher, after first separating the total record into subunits associated with reading and mathematics. Thus, in actual fact, the prorating factor was the total number of observations during either reading or mathematics since all observation variables were accumulated separately for these activities. These prorated frequency measures provided us with classroom variables which, we believe, can be considered independent, both of number of days of observations and of the possibly confounding classroom effects detailed above.

To study the relationship of teaching performances to reading and mathematics achievement in grades two and five, each APPLE variable was correlated with spring test scores (controlling for fall scores on each of the reading and mathematics achievement tests).

Other analyses from the APPLE substudy yielded data relating to the ways in which teachers distributed their classroom instructional time among the different contexts specified in Table 12. These time-allocation measures were derived from reports which required the observer to record for each pupil a maximum of the three most frequent contexts in reading (r_1, r_2, r_3) and the three most frequent contexts in mathematics (m_1, m_2, m_3) in which the pupil had been observed during the specified observation period. Reported for each of these contexts for each pupil were the following:

A = Total number of minutes in each of three contexts for each pupil

B = Percent of total time devoted to each context for each pupil, for example:

$$B_{r_1} = \frac{A_{r_1}}{A_{r_1} + A_{r_2} + A_{r_3}}$$

C = Total number of minutes in each context during which pupil was engaged in productive, student-like behavior

D = Percent of time in each context during which pupil was engaged in productive, student-like behavior, for example:

$$D_{r_1} = \frac{C_{r_1}}{A_{r_1}} \qquad D_{r_2} = \frac{C_{r_2}}{A_{r_2}}$$

For each of the four variables, scores for each teacher were obtained by averaging across observation days and across pupils in order to obtain four summary classroom measures, one reporting "average number of minutes" in each context, another reporting the "averaged proportion of the total time" in each of the contexts, and an additional two indicating the "average number of minutes of productive time" and the "average proportion of time which was productive" in each context.

In the review of the time allocation data, it became apparent that percent of time and percent of productive time measures would give only a partial picture of the total classroom instructional environment. If the amount of time the target pupil spent in each of his three primary contexts was not equivalent, or if all target pupils were assigned to the same contexts and spent different amounts of time in each, then it was felt that an additional method for aggregating the time allocation measures by teacher was needed.

This method yielded two supplementary measures. The first was obtained by adding the total number of minutes spent in each context across pupils for each teacher and dividing that figure by the total minutes in all contexts. This variable was called the "proportion of gross time" in context. The second measure was computed by adding, across target pupils for each teacher, the total number of minutes judged productive in each context divided by the total number of minutes across all target pupils in that context. This variable was called "proportion of gross productive time."

An illustration shows how "average proportion productive time" and "proportion gross productive time" can differ for data from the same teacher:

Total Time in Context	Total Productive Minutes	Percent Productive Minutes
T_1 = 5 minutes	1 minute	20%
T_2 = 15 minutes	5 minutes	33%
T_3 = 25 minutes	25 minutes	100%
		Average Pupil Percent Productive = 51%
Total Gross Time = 45 minutes	Total Gross Productive Time = 31 minutes	Gross Percent Productive = 69%

In this example, the gross percent productive time was greater (69%) than the average productive time across the target pupils (51%). As can be seen from the example, when pupils are more productive in contexts that have a longer duration, the gross percent productive time in context will be greater than the average percent of productive pupil minutes.

The choice of which value most fairly measures the classroom instructional impact will require further evaluation. However, for clarity and completeness of data presentation, we analyzed each of the following time allocation variables in its relationship to reading and mathematics performance and attitudes:

1. Average number of minutes in each context
2. Average proportion of total minutes in each context
3. Average number of productive minutes in each context
4. Average proportion productive minutes in each context
5. Average proportion gross number of minutes in each context
6. Average proportion gross productive minutes in each context

Description of Teacher and Pupil Behaviors. From the discussion of the statistics describing central tendencies and variabilities of these data across classrooms, instructional areas, and grade levels, several statements summarize the pupil and teacher behavior in the classrooms.

1. Differences in incidence of pupil events are a function of the instructional area, the instructional context and the teaching activities. This finding means that one must expect to find different types of pupil behavior and differing frequencies of pupil behavior, depending on whether the child is a second or a fifth grader, whether the child is involved in reading or mathematics lessons, whether the child is working with the teacher or independently, and what the activity of the teacher is.

2. In general, we can conclude that fifth graders will be expected to spend more of their classroom time working independently of the teacher for both reading and mathematics activities than will second graders.

While there is considerable variability among second-grade teachers in the use of different instructional contexts, second-grade pupils will be involved in more teacher-directed instruction than fifth graders.

3. The quality of teacher responses to pupil behavior is very similar across both grade and instructional area. The response of "None" to observed pupil behavior occurred about two-thirds of the time. We do not mean to imply that a teacher response should accompany every observed pupil behavior; rather, the data indicate the probability of the occurrence of any teacher response to pupil behavior in the classroom.

4. In organized instructional periods such as reading and mathematics, conduct problems and need for teacher disciplining are not common. This finding suggests that supervisors, and others who visit classrooms, cannot rely on the absence of conduct problems as good evidence for teacher competency. Inappropriate behavior during academic lessons was relatively rare.

The Allocation of Classroom Instructional Time. When we examined teacher performance with respect to the ways teachers distributed classroom instructional time in different contexts, these generalizations followed from the descriptive statistics.

1. Second graders spend more total minutes in reading instruction than do fifth graders, and this is primarily reflected in more time spent in group instruction and time spent being taught by other adults. More minutes were spent in independent-individual contexts, those contexts in which pupils are working on an individualized assignment different from other pupils and independently of the teacher, in second-grade reading than in second-grade mathematics and also in fifth-grade reading instruction.

2. One surprising finding was that the average proportion of pupil productive time in individual contexts in reading was relatively small in both second and fifth grade. Few pupils receive such instruction and of those who did, the average percent of productivity as judged by the observers was lower than the percent of gross productivity in all contexts. This finding implies that the teacher's choice of this context for certain pupils is not always a wise one. Only those pupils who can manage such a context for significant periods of time (probably greater than 5 to 10 minutes) are likely to profit from it at all.

3. The primary difference in time allocation in mathematics between grades was the relatively greater proportion of time spent in Teacher-Class contexts at second grade.

4. The time spent in Independent-Individual, Adult-Individual, Teacher-Individual and Adult-Group contexts is not generally productive time in mathematics when the percent of productive time was averaged across pupils. However, when the gross productive time was computed by finding the proportion of productive time by classroom, rather than averaging across pupils, the productivity rate for these contexts was as high as for other contexts. This finding, along with the similar result reported for reading, *indicates that only a few pupils work productively in individual contexts*. Therefore, on the average, pupil productivity is lower in individual contexts than it is in settings where instruction is by group or by class. However, individual contexts are appropriate and productive for some pupils. Thus, when one looks at the rate obtained by dividing total productive time for all target pupils by total time in that context for all target pupils, the productivity is higher.

Validity of Observational Variables for Predicting Reading Outcomes. The great variability in pupil and teacher behavior between grade levels and between instructional areas leads to the preliminary conclusion that we cannot expect to find a set of teacher performance variables which will be valid for predicting learning outcomes across grade levels and instructional areas. The results of the partial correlations between observation variables and the reading achievement measures support this initial impression. Some of the generalizations that we can make on the basis of correlations between observation variables and reading achievement are:

1. Talking out inappropriately in classrooms during reading instruction was uniformly related to poorer reading outcomes. There was also an inverse relationship between the observed frequency of pupil behavior which was not consistent with teacher expectations (but not necessarily serious enough to warrant teacher intervention) and achievement outcomes.

2. Evidence of positive achievement in phonics skills was positively correlated with outcomes for second grade and negatively correlated with reading outcomes for fifth grade. This result suggests that teacher instruction in phonics skills enabled second graders to improve their achievement. However, it appears that teacher instruction in phonics at fifth grade was directed at pupils with learning difficulties and perhaps was not the most effective means by which to improve learning for all pupils in the classroom.

3. Conversely, evidence of positive achievement in oral reading was positively correlated with outcomes at fifth grade, but not at second grade where oral reading is regularly observed for all pupils in reading instruction.

4. With regard to teacher activities, we offer the tentative hypothesis that the more the teacher is involved in organizing the class for instruction (as evidenced by negative correlations between activities like "Academic Organizing," "Asking Questions," "Discussion," and "Instruction Giving," and reading achievement) and the less the teacher is involved in "Working With" pupils and "Checking" to see if they understand their work, the poorer the end of the year reading achievement. We do not intend to imply that teacher activities like "Academic Organizing," "Asking Questions" and so forth are not important teacher performances. However, as the frequency of these activities increases, the frequency of actual instruction decreases. Consequently there will be reduced opportunity for reading instruction, resulting in less improvement in reading at the end of the year.

Validity of the Observational Data for Predicting Mathematics Outcomes. In comparison to reading outcomes, a different set of observation variables predicted mathematics learning. These conclusions follow from the analysis of the predictive validity of the observation variables:

1. In second grade mathematics lessons during which there is a high frequency of inappropriate conduct, obvious inattention to classroom work, and talking out of turn, there is more likelihood that achievement scores in mathematics will not increase by the end of the year.

2. Though noticeably inappropriate behavior did not characterize fifth-grade pupils in mathematics assignments, the frequency of inattention to task which was not disruptive or particularly noticeable was uniformly negatively correlated with mathematics learning.

3. The frequency of Independent-Group and Independent-Class contexts was negatively correlated with achievement outcomes at both grade levels, suggesting that classrooms in which pupils were working independently of teachers most of the time were those with poorer mathematics achievement.

4. A surprising finding was that teacher instructional activities were observed with less frequency in mathematics than in reading. Only "Asking" had a significant correlation with outcome in mathematics and it was negatively correlated with achievement. The tentative conclusion is that teachers who spend time asking questions are those whose lessons are not understood or whose assignments are at an inappropriate level of difficulty.

5. Three out of 12 teacher responses at second grade and five of 12 at fifth grade had a significant relationship to mathematics achievement. At second grade, teachers who employ "disciplining and recognizing," have pupils with pupils with poorer than expected achievement. At fifth grade, "giving help," "providing positive feedback," and "giving praise" are positively correlated with pupil learning. "Recognizing" and "no response" are negatively correlated with mathematics learning at fifth grade.

Validity of Observational Variables for Predicting Attitude Outcomes. A most interesting finding was that teacher performance variables which predict achievement may have an inverse relationship to changes in children's attitudes toward learning and toward themselves.

1. Different sets of appropriate and inappropriate pupil behavior were correlated with reading and mathematics attitudes both in second and in fifth grade. We have to conclude, therefore, that it is unrealistic to assume that one can define an omnibus set of teacher performance variables, which, if present in a group of teachers, will guarantee positive learning outcomes as well as positive attitudes.

2. With respect to attitudes toward reading and mathematics, second-grade classrooms in which target pupils were observed to be positively engaged in the activities of the class and in which they were not misbehaving or talking out inappropriately were those with the most positive attitudes at the end of the year.

3. Since the Pupil Survey items ask questions about how the children judge their abilities with respect to peers, it was not surprising that some pupil event variables related to appropriate pupil behavior had negative correlations with the survey measures. In some classrooms, positive participation of some pupils may cause other pupils to contrast the adequacy of their performance unfavorably with that of their classmates, resulting in a negative correlation between some positive behaviors and the survey measure.

4. At second grade, "Academic Performance," "Attentiveness" and "Direction Following" had positive relationships to the survey measure, suggesting an age difference in these types of pupil behavior on attitudes toward self-worth.

5. There seemed to be no significant relationship between the incidence of the instructional contexts and attitudes across either grade level or subject. Attitudes, therefore, are more dependent on teacher and pupil behavior in the classroom than on the subject.

6. The teacher activities which are significantly related to attitude and are always positively correlated with attitude and survey outcomes across instructional area are "Checking," "Listening" and "Supervising." Those variables which have uniformly significant negative correlations are "Circulating" and "Instruction Giving."

7. Teacher responses to pupil behavior had different patterns of relationships to attitudes according to whether teacher responses were observed in reading or mathematics and whether pupils were second or fifth graders. The data indicated that the incidence of disciplining and negative feedback in a classroom cannot be judged to be always negatively correlated to general attitudes in the classroom. On the other hand, the extent of positive feedback was always positively related to attitudes when the partial correlations were significant.

8. From the analysis of teacher activities and responses, we tentatively infer that teacher behavior which calls attention to children, and perhaps places them in an unfavorable comparison with their classmates, may have a negative effect on attitude, even though these same teacher behaviors have a positive relationship to achievement outcomes.

There are enough differences between grade levels and between reading and mathematics instruction that we cannot generalize across grades and instructional areas regarding which teacher performances will always predict pupil achievement or attitude outcomes.

Time Allocation Data as Predictors of Learning and Attitude Outcomes. The length of time pupils spent in some instructional contexts relates to achievement and attitude outcomes. Some of the findings were:

1. The more time pupils spent in Independent-Individual contexts in second-grade reading, the better their learning in reading comprehension

2. While the amount of time second-grade pupils worked independently and productively in individual contexts was correlated with the reading comprehension measures, the amount of time the teacher spent conducting reading lessons for the class was significantly correlated with the Decoding

Test scores. The data offered clear evidence that the learning of decoding skills at second grade was dependent upon the amount of learning time during which the teacher is in charge of class instruction.

3. The amount of learning time in context does not always have a positive relationship to outcomes. An example was the negative correlation between the Independent-Group contexts and the reading measures. Time measures may have a curvilinear relationship to achievement. Some efforts to determine the amount of time which is the optimum for learning in each context would be an excellent way to extend our understanding of the relationship between the duration of classroom experiences and learning.

4. Fifth graders who spent more time in contexts where teachers are in charge of instruction have higher scores on reading comprehension measures.

5. For second graders the most effective context time-allocations for mathematics learning were the several individual contexts. Even though individual contexts were common to only a few BTES classrooms, the data indicated that the more time second graders spent in individual contexts, and the more this time was productively spent, the better their mathematics learning.

6. While time spent in individual contexts in mathematics was important to second-grade mathematics achievement, the amount of time pupils were taught mathematics by the teacher in fifth-grade classes was significantly related to outcomes.

7. In fifth grade both mathematics and decoding skills require the learning of rules and principles and practice in applying these rules and principles. The more time the teacher spent directing these learning experiences, the better the learning outcomes.

The results of our preliminary examination of the relationship of these variables to student learning and attitude showed that teacher performance does make a difference when the effects of the initial pupil learning and attitude are controlled.

Of considerable interest, and worthy of additional study, was the finding that using the classroom as the unit of analysis, APPLE observation variables which were significantly associated with reading and mathematics achievement were not necessarily those associated with positive attitude change. The fact that teacher performance variables which were related to significant increases in the criterion measures were specific to those measures indicates that we must study teacher performance with multidimensional outcome criteria.

Teacher Performances Associated with Different Kinds of Teaching Effectiveness. A major use of the APPLE observation data in the BTES study was to treat sets of observation variables as though they were items in a category observational system. Additionally, the system was used to identify pairs of observation variables across categories. For example, one might be interested in the types of pupil behavior which occurred during different types of teaching activities, or the types of teacher responses associated with different types of pupil behavior. There are numerous interesting hypotheses relating to the dependence of pupil learning on combinations of teacher-pupil behaviors or instructional context-teaching activity behaviors.

For example, one of our underlying assumptions about teacher performance characteristics in elementary education programs is that the way the teacher organizes instruction determines the extent to which pupils in the classroom are under the teacher's direction or function independently of the teacher's instruction. We believe that the teacher's organizational pattern can affect learning outcomes. To provide a preliminary test of this assumption using the data of the BTES study we identified at each grade level and in each subject area the five teachers whose outcomes for pupil achievement were better and the five teachers whose pupil achievement outcomes were poorer than would have been predicted by fall test scores. The example presented here was selected from a variety of such analyses to illustrate the differences in teaching performance between these two groups of teachers.

The five top and five bottom teachers were located by performing regression analyses for a particular achievement outcome. When these analyses were studied, we found differences in the pattern of gains made during the year. These differences in achievement effects could be organized into classes where all students made gains from fall to spring (top) and into classes where only those students who were high on the fall test made gains (bottom).

Table 13 contains a tabulation for the top and bottom groups of the frequency of teaching activities within instructional contexts for second grade reading. (We have provided definitions of selected instructional contexts and the selected teaching activities in Table 12.)

The table is organized in the following ways. The column percents under each teacher activity provides information as to how top and bottom teachers differ with respect to the frequency of instructional contexts associated with a particular teaching activity. Note that for the top teachers, two contexts were observed when the "Circulating" activity was also seen (I-Class in 85.7 percent of the instances and I-Individual in the remaining 14.3 percent). For the bottom teachers, "Circulating" never occurred, regardless of context.

Table 13

Frequencies of Teacher Activities in Contexts for Teachers with Better than Predicted and Teachers with Poorer than Predicted Outcomes in Second Grade Reading

Contexts		Academic Organizing		Answering		Asking		At Board		At Desk		Checking		Circulating		Discussing		Explaining		Helping		Instruction Giving		Listening		Question & Answer		Supervising		Work With		Percent Row Total
		Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	Col %	Row %	
A-Group	Top	16.1	11.1																			6.7	2.2							11.9	86.7	8.2
	Bottom	7.7	11.1			33.3	22.2																							4.7	66.7	5.1
I-Class	Top	6.5	7.4			5.0	11.1					22.9	29.6	85.7	22.2					40.0	7.4	6.7	3.7					25.0	3.7	1.2	14.8	5.0
	Bottom	23.1	15.0	100.0	35.0															37.5	30.0									3.1	20.0	11.4
I-Group	Top	41.9	7.9			1.7	.6					31.4	6.7					20.0	1.2			6.7	.6	21.1	2.4					40.6	80.6	30.8
	Bottom	7.7	1.3			16.7	1.3																							59.8	87.4	44.3
I- Individual	Top	3.2	4.0									17.1	24.0	14.3	4.0															5.2	68.0	4.6
	Bottom																													2.4	100.0	3.7
T- Class	Top	19.4	14.6			25.0	36.6					2.9	2.4			11.1	2.4	30.0	7.3			13.3	4.9	5.3	2.4	7.1	2.4	50.0	4.9	2.7	22.0	7.6
	Bottom					16.7	6.7													43.8	46.7	75.0	20.0					100.0	6.7	2.4	20.0	8.5
T-Group	Top	12.9	1.7			68.3	17.3	100.0	.8			25.7	3.8			88.9	1.4	50.0	2.1	60.0	1.3	66.7	4.2	73.7	5.9	92.9	5.5	25.0	2.4	38.6	53.6	43.9
	Bottom	61.5	15.7			33.3	3.9													18.8	5.9	25.0	2.0	10.5	2.0	100.0	2.0			27.8	68.6	22.2
Percent Column Totals	Top	5.7				11.1		.4				6.5		1.3		1.7		1.9		.9		2.8		3.5		2.6		.7		60.9		
	Bottom	7.4		4.0		3.4														9.1		2.3		.8		.6		.6		72.2		

Similarly, reference to the row percents shows how the top and bottom teachers differ with respect to the frequency of activities observed within a particular context. Consider "A-Group": in the classrooms of the top teachers, 8.2 percent of the observations were of children working in a group with an adult, and 5.1 percent of the observations in the bottom teachers' classes were in this context. The examination of the teaching activities that were taking place, however, shows that "Instruction Giving" by an adult other than the teacher was observed only in the top teachers' classes.

One interesting question is whether activities in different contexts vary between the top and bottom teachers. Table 13 illustrates the important finding that the top teachers used a greater variety of teaching activities than the bottom teachers, and that there were differences in the relative frequencies of these activities. Overall, the "Work With" activity was the activity most frequently observed for both groups of teachers, but the bottom teachers were observed relatively more frequently working with pupils than were the top teachers. Teachers in the bottom group also were observed helping more, spending more time organizing the lesson and answering questions more often. Teachers in the top group were observed more frequently asking questions, checking pupils' work, giving instruction, listening, and conducting question-and-answer sessions.

Using these data, it may be hypothesized that in second-grade reading not only do more effective teachers engage in certain teaching activities more frequently but they also employ a wider variety of activities. They conduct instruction by actively interacting with the children, checking the work, asking questions, explaining, giving instructions, and checking for understanding with question and answer sessions.

Summary. A behavior recording system such as APPLE is a useful tool for studying teacher performance characteristics. The APPLE system makes few *a priori* assumptions about the nature of teacher and pupil behavior, can describe pupil and teacher characteristics with considerable fidelity, and identifies those pupil and teacher behaviors which are significantly related to learning for a particular grade and instructional area. From the use of a naturalistic observation system like APPLE, we will be able to deduce a theory of instruction which is based on actually occurring teacher behaviors rather than theories which are dependent on laboratory studies with experimentally manipulated teacher variables.

The RAMOS Observational System

The RAMOS[16] Observational System was one of two observational systems used in the study. The RAMOS System is a time-based category system of observation. "Scores" from the observations were used as measures of teaching performance for the analyses done in the study.

The Reading and Mathematics Observation System, RAMOS, is designed for real-time documentation of classroom activities of teachers, student groups, or individual students. It is best suited to observation of classroom instruction in the areas of reading and mathematics at the elementary level. With it, a trained observer can record continuously the events in a regular classroom in comprehensive detail and in real time. The system permits the observer to focus either on the teacher, a group of students, or a small number of target students, depending on the purpose of the observation.

---

[16] Reading and Mathematics Observation System: Description and Analysis of Time Expenditures, Robert Calfee and Kathryn Hoover Calfee, Volume III, Chapter 2, Phase II Final Report.

Once an observation interval (generally between 30 minutes and an hour) is initiated, the observer records the events that transpire as a series of "line," each line containing categorical entries describing the nature of the event. These lines provide answers to such questions as:

> Who are the students being observed? How many are there in the group? Where are they located in the room? Which target students, if any, are in the group?
>
> Who are the adults in the classroom? How many adults are there? Where are they, and with what student groups are they associated?
>
> What is going on at any given time? What is each student group and adult doing? If engaged in an instructional activity, how many students and adults are involved, and where are they?
>
> Who is doing the instructing? What is the nature of the instructor (teacher, aide, volunteer, tutor, etc.)? What role is the instructor playing (direct instruction, discipline, class management, etc.)?
>
> What is the content of instruction? What subject matter is being taught? What skills and activities are involved? What materials are being used? What kinds of feedback are available to the students?
>
> What is the response of the group to instruction? What kinds of responses are required or expected of the students? What is the judged level of attention to instruction? Relative adequacy of performance? Amount of social interaction between students?

From the records obtained in this manner, several measures were extracted for the purposes of the BTES project. The primary aim of these measures was to describe the manner in which time is used in the classroom. The measures included (1) total time spent on activities directly or indirectly related to reading and mathematics, (2) the character and variety of those activities, and (3) the relative distribution of time spent in various activities.

Procedures for Conducting a RAMOS Observation. A RAMOS observer used several forms while observing in Phase II classrooms: a Start Form, an Event Form, a Debriefing Form, a Classroom Summary Record, and General Descriptive comments. The Start Form and the Event Form were basic to the results described in this paper.

A RAMOS observer begins the observation by filling in a Start Form, a labeled computer coding form for recording the teacher, school, district, and physical characteristics of the classroom about to be observed. Characteristics of the student groups and the instructors are also recorded before the actual observation begins.

The Event Form is a labeled coding form used for recording each classroom event or change in real time. The initial condition of all existing groups is fully described, the instructor and the content of instruction is documented, and certain aspects of the student response are recorded. After the first lines have been completed, further lines are added as necessary to document any discernible changes in a group or in a target student within a group. Movement of students between groups or the formation of a new group also requires a line.

Each event line records the time an event begins and indicates the group involved. Notations specific to instructors or target students are coded with the group association. Five major categories are recorded on each event line in RAMOS: Time, Group, Instructor, Content, and Response. Within each category, codes are entered in fields from the RAMOS Summary Codes sheet (Figure 3).

RAMOS

SUMMARY CODES

Time:
24 Hr. clock

0000	Start file
0001	Gener info. line
0002	Group ID line
0003	Instructor ID line
0004	Continuation line
0005	General comment
0006	to
0050	Field-specific com.
0099	End file

Group
Status:

G	Group
T	Teacher only
P	Part of a group
S	Specific students
B	Background only
N	Not being observed

Group ID:
from START

[illegible]
0-99

Grid:
code from START

Location:

S	Students seats
T	Table for group work
C	Circle grouping
G	Game Area
D	Desk of teacher
L	Library area
B	Blackboard
N	Not fixed or moving

Density:

D	Dispersed singly or in twos
S	Small groups
C	Cluster of majority
W	Whole class together

INSTRUCTOR

12 **Instructor ID**
From Start

13 **Classifications:**

T	Teacher
R	Resource teacher
N	Intern teacher
A	aide
V	Volunteer
C	Crossage tutor

14 **Role:**

N	Instruction
F	Facilitation
D	Discipline
M	Manage class
A	Assess/diagnose
E	Evaluate for mastery
P	Preparation
S	Supervise staff
I	Independent

15 **Availability:**

A	Available
N	Not available
E	Emergency
S	Specific student
G	Group being taught

16 **Interaction:**

T	Instructor chooses
V	Volunteers chosen by instructor
H	Help sought by students
B	Both student and instructor initiate
R	Routine of turns

17 **Mobility:**

M	Mobile
L	Limited
S	Stationary

CONTENT

18 **Subject:**

R	Reading
L	Language
M	Math
C	Science
S	Social Studies
A	Art/music
B	Business of class
F	Free choice
W	Wait time
P	Playtime, recess

19-22 **Reading Skills:**

DS	Decoding, simple
DA	Decoding, adv.
SS	Structural analysis, simple
SA	Structural analysis, adv.
SY	Syllabification
WM	Word meaning
DW	Dictionary work
RA	Reading aloud
RS	Assign. silent reading
RR	Recreational reading
CF	Compre. lit facts
CM	" main ideas
CR	" relations
CE	" external. infer.
CS	" sequence
CA	" apprec.
CC	" conclu.
GR	Grammar
PU	Punctuation
PS	Practice spelling
PP	Penmanship
WA	Assigned writing
WC	Creative writing

23-26 **Math Skills:**

B_	Basic facts
S_	simple
A_	Advanced
F_	Fraction
D_	Decimal
%_	Percent, ratio, prob.
W_	Word problem
_+	Addition
_-	Subtraction
_*	Multiplication
_/	Division
_R	Relationships
RS	Set relations
GR	Graphing
GE	Geometry
ME	Measurement
TM	Time
$$	Money

27-28 **Materials:**

B	Book
C	Chart/chalkboard
W	Workbook
M	Manipulative
P	PaperPencil
E	Equipment

29-30 **Aim:**

N_	New
P_	Practice
R_	Review
A_	Application
_C	Concept
_S	Skill
_F	Fact

31 **Activity:**

L	Lecture
D	Discussion
S	Seatwork
R	Recital - reading res
Q	Question-answer
A	Audio Visual
G	Game
V	Visual demonstra.

Feedback

32 **Pattern:**

D	Delayed verbal
M	Immediate verbal
E	Expression
P	Physical contact
T	Token
W	Written

33-34 **Feedback Sign:**

+_	Positive
-_	Negative
B_	Both
_U	Undifferentiated
_T	Task specific
_B	Both

Group Response

35 **Task:**

L	Listen
S	Speak
R	Read silently
A	Read aloud
W	Write
P	Physical act

36 **Focus:**

N	Instructor
C	Classmates
L	Learning material
S	Something else

37 **Social Interaction:**
38 **Contact with instructor**
39 **Output rate**
40 **Performance**
41 **Attention/involvement**
42 **Physical activity**
43 **Noise**

H	High
M	Medium
L	Low
X	None

Special codes:

X	Stop code
?	Uncertain

R. Calfee & K. Hoover
Stanford, California
Experimental version, not to be reproduced or used without explicit permission.

Figure 3

Copyright © 1974 by Robert C. Calfee and Kathryn A. Hoover. All Rights Reserved.

The fields for the TIME category are the time an event begins, recorded to the nearest minute on a 24-hour clock, and status, an indication of the focus of the observer at this time.

The GROUP category fields are the ID number indicating to which group the line applies, the number of students in the group, grid or relative position within space, location within room, and density of students in the group.

The INSTRUCTOR category describes the activities of the instructors at a given time by the following fields: instructor ID shows the specific instructor, followed by his classification, the role he is now assuming, his availability to students, how student-instructor interactions are being conducted, and mobility of the instructor.

The CONTENT category fields include the subject being studied, reading main and second skill, math main and second skill, main and secondary material, aim of instruction, activity used for instruction, feedback pattern of instructor, and feedback sign to students.

The RESPONSE category fields indicate the task assigned to the group of students, the focus of attention of the students, their level of social interaction, output rate, quality of performance, level of attention and involvement, physical activity, and noise level. There is also a column to record a code indicating when a specific target student is referred to by this line.

Editing and Analysis of RAMOS Protocols. The basic unit of analysis for the RAMOS protocols is the student minute--a single-number minute of student time during a class period. The RAMOS analysis program computed, within each category of the system, the number of student minutes falling within each of the codes assigned to that category. For instance, suppose that, during a given episode, one group of ten students spends 20 minutes reading, 5 minutes in class business; a second group of six students spends 12 minutes in mathematics, 8 minutes in reading and 5 minutes in class business. Then under the SUBJECT category, the distribution of student minutes would be 248 for the Reading code (20 x 10 for the first group plus 6 x 8 for the second), 72 for the Mathematics code (6 x 12 for the second group), and 80 for the Business code (10 x 5 for the first group plus 6 x 5 for the second group). Of the total 400 minutes in the episode, the Reading code has the highest percentage, 62 percent, followed in second place by Business with 20 percent and Mathematics with 18 percent. The analysis program computed the quantities -- student minutes and percentages -- for each code, and arranged the results in rank order from highest to lowest. The series of observations for each of the 95 teachers were then combined into a single record for each teacher-class for use in various analyses. (Each teacher-class was observed on one to four occasions.)

Results. The primary interest in this study was to uncover any strong relations of teacher characteristics and practices to student outcomes in reading and mathematics. Accordingly, time usage in the areas of reading (and closely allied language arts) and mathematics was chosen for analysis. Within these subject matter areas, four major categories were selected for inclusion in the analysis: role of the instructor, aim of instruction, instructional activities, and main materials. These categories answer the questions: What was the teacher doing (Role), to what purpose (Aims), by what means (Activities), and with what resources (Materials)? These categories were chosen a priori as most likely to bear a relation to the overall pupil measures of reading and mathematics.

The relative distribution of time usage by the target teacher (the "main instructor") in various roles showed considerable variation in time spent in direct instruction and facilitation as a function of grade level and subject matter. Direct instruction includes situations in which the instructor is maintaining direct control over learning activity. Facilitation refers to activities in which the students work independently, but the instructor provides advice and feedback at the student's request or as a result of the instructor's direct observation.

The results indicated that reading and language arts were much more likely to proceed by direct instruction, whereas mathematics was more likely to be under the pupil's control with the teacher providing help when the pupil asks for it or when the teacher sees something that requires attention. Direct instruction is far more common in second grade than fifth grade, in both reading and language arts and in mathematics.

Management, evaluation, and other activities occupied about 25 percent of the teacher's time under all conditions. Assessment and evaluation are more likely to take up time in observable ways in second grade than in fifth grade.

The second major dimension is related to the aim of instruction. The dominant aim, the aim which accounted for most of the time, is practicing skills, either directly or through review of some sort. Relatively little time was spent on new ideas or topics, and only in fifth-grade language arts was any substantial amount of time spent in the application of knowledge, skills or concepts to practical problems beyond a school context.

The breakdown of time usage by grade and subject matter for the category of instructional activities suggested that students spent most of their time in seatwork. This activity was more common among fifth graders than second graders, and was more common in mathematics than in reading. This pattern parallels the variation in faciliation as an instructional role. Discussions and question-answer episodes took up about 20 to 25 percent of the time. About 10 percent of the reading time in second grade was spent in oral reading recitation and about the same proportion of the mathematics time was spent in games of some sort.

Analysis of time usage for the category "Main Materials" revealed large shifts in patterns of materials usage which varied as a function of grade level and subject matter. Main instructional materials were those that seemed of chief importance to the activity carried out by the pupil. Books were used more in reading than in mathematics, and more in fifth than in second grade -- an expected result. Workbooks were common everywhere, but especially in second-grade mathematics. They also occupied a surprising amount of the time in fifth-grade language arts and reading. About 15 to 20 percent of the time, pupils were left on their own with paper and pencil. In second grade, pupils were working with no materials about 20 percent of the time; it seems likely that much of this time was accounted for by class discussion. Only in second-grade mathematics was a considerable amount of time spent with manipulatives (Cuisenaire rods, number scales, etc.) or equipment (film strip projectors, tape recorders, etc.). Generally, instruction in these classrooms was carried out with the old standbys -- books, paper and pencils, and chalkboard.

If other materials also played a significant role, this was recorded as secondary material. Secondary materials reflect the extensiveness of the instructional materials being used at the time in the classroom. For example, in mathematics instruction, it is the difference between a lesson in a workbook and a lesson in which the student also uses a number scale to check his work. An example in reading is the difference between a lesson where the student reads to himself from a book and one in which he records his thoughts and reactions in writing as he reads. Secondary material usage was greater at fifth grade (55 percent) than at second grade (45 percent), and was greater in mathematics (55 percent) than in reading (45 percent). Paper-and-pencil was the most common secondary material.

Two sets of correlations among various combinations of the RAMOS variables were especially noteworthy. One seems to encompass a pattern of direct instruction in which books and workbooks are used by students at their seats for practice and review purposes. The second pattern appears to represent situations in which teachers are more likely to spend their time in something other than direct instruction, in which students are engaged with new concepts and facts with unconventional materials or group discussion. More detailed examination of the data is needed to confirm the character of these patterns but they do seem to parallel descriptions of contrasting teaching styles found in contemporary classrooms.

These are the variables that were entered into the structural analysis. They represent aggregates over complex patterns of activities, and undoubtedly some useful relationships have been obscured along the way. However, given the thesis that total time usage is one of the more fundamental variables determining instructional effectiveness, these summary variables represent a reasonable choice for initial stages of hypotheses generation.

The Reports of Teachers About Their Instructional Activities

Teacher reports[17] of their instructional activities and procedures were used as one source of descriptive information about classrooms. Work Diaries and Program Overviews were one of several approaches used to describe teaching performances in the study. The instruments were intended to provide a vehicle for teachers to describe their programs in reading and mathematics directly. We were interested in whether teacher reports would be accurate and valid sources of information about classrooms. Two types of teaching performance variables were derived from direct observations in classrooms. The third type of teaching performance variable was derived from the Work Diary scores.

Scores were aggregated from the four instruments teachers used to describe their instructional activities. The "Reading Program Overview" and the "Mathematics Program Overview" were intended to provide a summary of the teacher's program for the entire school year. These two instruments were completed once by all 95 teachers. The "Reading Work Diary" and the "Mathematics Work Diary" were designed to reflect the instructional process during one school week. The Diaries were completed during two target weeks which were eight calendar weeks apart. All of the 95 teachers returned the first set of Diary materials, and 93 returned the second set.

---

[17] The Reports of Teachers About Their Mathematics and Reading Instructional Activities, Patricia Elias, Gail Hare, and Patricia Wheeler, Volume V, Chapter 5, Phase II Final Report.

Because the instruments were intended for teachers to use with little or no direct project staff support, it was necessary that both the structure and content be understandable and sensible to the teachers. Review by project teachers was an integral part of developing the Diaries.

The information from the Work Diaries and Program Overviews was organized to indicate, by grade level and within subject areas, the academic emphasis, the quality of the teaching methodology, the complexity of organizational structure, and the variety and types of instructional materials used. The accuracy of sections of the teacher reports was tested against observational information from the project classrooms and was related to pupil skill growth in reading and mathematics.

Academic Emphases. Academic emphases included three general information categories: how much time teacher reported spending in preparing for and in teaching reading and mathematics skills, how frequently teachers reported teaching specific reading and mathematics skills, and the specific skills in reading and mathematics that teachers reported they did not plan to teach. Some of the more interesting results of how teachers reported spending their time are presented below.

1. Second-grade teachers spent more time preparing for and teaching reading than did fifth-grade teachers.

2. Fifth-grade teachers spent more time preparing for and teaching mathematics than did second-grade teachers.

3. The amount of total teacher preparation time had no relation to pupil skill-growth in mathematics and had an inverse relation to pupil skill-growth in reading. Greater amounts of preparation time were associated with poorer reading comprehension skills for fifth-grade pupils.

4. The amount of direct teaching time had no relation to pupil skill-growth in mathematics but did relate to second-grade pupil skill-growth in decoding.

5. There was greater grade-level variation in time spent teaching specific reading skills than there was in the teaching of specific mathematics skills.

6. Second-grade teachers, as would be expected, spent more absolute and proportional time teaching decoding and vocabulary skills than did fifth-grade teachers.

7. Both second and fifth-grade teachers spent more time teaching more generic reading comprehension and reading application skills than they did teaching more specific decoding and vocabulary skills.

8. Greater amounts of time spent teaching the more specific decoding and vocabulary reading skills to fifth-grade pupils were associated with poorer pupil performance on the more generic reading tasks, comprehension, application and achievement.

9. The time second-grade teachers reported spent in teaching specialized skills in mathematics was negatively related to pupil growth in computation skills, and positively related to concept skill-growth.

The time teachers spent teaching specific reading and mathematics skills in relation to pupil performance suggests a variety of research tasks which might focus on the identification of optimal ranges of time for teaching specific skills to pupils in classrooms with a variety of skills configurations. Our data suggest, for example, that fifth-grade teachers who emphasize basic decoding and vocabulary skills during reading instruction may do so at the expense of pupil skill-growth in comprehension skills. (The time teachers reported spending in preparing for and in teaching reading or mathematics was the basis for the combined category, WD-1, included in the analyses described earlier in this summary.)

Teachers were asked to check skills that they had "introduced" or "reviewed" during each of the two weeks the Diaries were maintained, in other words, skills they had actually taught. Teachers were also asked to indicate in the Program Overviews which specific skills would not be taught during the school year, either because "pupils already know these skills" or "these skills are taught at a higher level." (The frequency of skills taught and not taught provided the basis for the combined category score WD-2, included in the analyses described earlier in this report.) Some of the results of the reported frequency of skills taught were:

1. The number of skills taught had little proportional relation to the time spent teaching within a skill area. Second-grade teachers, for example, reported spending about twice as much time teaching decoding skills as did fifth-grade teachers. The numbers of actual skills taught, however, did not appear that different (ten skills and eight skills, respectively).

2. As was the case in time spent, there was greater grade-level variation in number of skills taught in reading than there was in mathematics. But where the number of skills taught differed, it made a difference in pupil skills, particularly in mathematics.

3. The number of specific skills taught related to pupil skill-growth in mathematics at both grade levels, and was particularly important for second-grade pupils.

4. The number of teachers who did not plan to teach specific skills was relatively small in both reading and mathematics. But the decision not to teach certain skills related to pupil outcomes in several instances. For example, the decision not to teach certain decoding and vocabulary skills by a relatively small proportion of second-grade teachers was related to poorer decoding and vocabulary skill-growth of their pupils.

The Quality of the Teaching Methodology. Teachers were asked to describe the teaching strategies, pupil activities, and materials used both for the skills they introduced and the skills they reviewed during a given week. Their responses were rated. (The "scores" became the combined category WD-3, included in the analyses described earlier in this report.) Neither score by itself related to pupil skill-growth at the second grade; ratings of descriptions of skills reviewed related negatively to pupil skill-growth in fifth-grade reading.

The Complexity of the Organizational Structure in the Classroom. Organizational structure included two general types of information about the project classrooms. Teachers reported how they grouped children for reading and mathematics instruction, and what proportion of the pupils were taught reading or mathematics only by themselves. The ratio of teaching adults to children was also determined for both subject areas. (These variables became the basis for the combined category of teaching performances called WD-4 earlier in this summary.) Some of the findings were:

1. Pupils were grouped differently for reading and mathematics instruction. Simultaneous instruction of all pupils was the most common pattern used for mathematics instruction, and grouping by general reading level was the most common for reading instruction at both grade levels.

2. Grouping practices did not relate to fifth-grade pupil achievement in mathematics. However, they did relate to several kinds of pupil achievement in reading and in second-grade mathematics. For example, grouping by general reading level, the most common convention, related positively to second-grade pupil skill-growth in application skills and negatively to decoding skill growth for fifth grade pupils.

3. Few teachers reported individualized instruction as their preferred organizational mode.

4. Individualized reading instruction was associated with poorer second-grade pupil performance on reading application tasks, but individualized mathematics instruction was associated with second-grade pupil growth in mathematics skills.

5. More fifth-grade pupils were taught reading and mathematics exclusively by their teacher than were second-grade pupils.

6. Second grade teachers utilized up to six teaching adults during reading instruction, and up to five during mathematics. Fifth-grade teachers who shared teaching or preparation responsibilities did so with a maximum of one or two other adults.

7. Pupil skill-growth in reading, particularly that of second-grade pupils, was associated with the proportion of pupils taught exclusively by their teacher. No such association was evident for mathematics instruction.

The generalization that the percentage of pupils taught exclusively by the teacher related to reading skill-growth at both grade levels was an intriguing finding. A variety of teaching configurations was represented in the sample: team teaching, multiple level, nongraded and open classrooms. A variety of human and school resources was available to some teachers: aides, student teachers, volunteers, reading specialists and learning laboratories. Under the circumstances (particularly in the case of the younger pupils) the simple generalization that pupil skill-growth in reading may relate to the proportion of pupils taught exclusively by the teacher should be studied further.

The Variety and Types of Instructional Materials Used in Reading and Mathematics Programs. Teachers were asked to approximate the percentage of their reading and mathematics programs in which they used the types of materials listed. (The materials used in reading and mathematics instruction were aggregated somewhat differently to provide the basis for the combined category WD-5 described earlier in this report; there the number of different kinds of materials used became the index for WD-5.) The teachers were provided with a list of twelve different types of reading materials and ten different types of mathematics materials. Space was provided to specify the exact materials used and to indicate any types not included in the list.

1. Patterns of reading-materials usage varied between second and fifth grade. Second-grade teachers, for example, used basal readers more than did the fifth-grade teachers, who relied more on individualized materials and reading kits.

2. There was little percentage difference between second and fifth-grade usage of the ten types of mathematics instructional materials.

3. The use of the state-adopted basal reading series was associated with second-grade pupil growth in application skills.

4. The use of the state-adopted basal mathematics series at fifth grade was associated with poor pupil performance in computation skills.

5. The use of programmed materials was associated with second-grade pupil growth in computation and total mathematics skills.

Teacher Reports of Their Activities and Observations of Those Activities. The validity of teacher reports of their activities was explored by comparing their reports to our observations of their teaching. Our purpose was to determine if reports provided by teachers related to other information obtained in the study. The comparisons made were classroom organization patterns and time spent teaching specific reading and mathematics skills as compared with observed evidence of positive pupil learning.

Teachers were asked to indicate in the Work Diaries which one organizational pattern most nearly described their current reading program. Their choices were: simultaneous instruction of all pupils, pupils grouped by reading level, pupils grouped by skill needs, individualized instruction and an open-ended "other" category. The four specific organizational patterns reported were compared with the six APPLE instructional contexts in which pupils were observed to be working.[18] Observed and teacher-reported organization of pupils for learning seem consistent to the degree that further research would be warranted to refine systems for teachers to report their own organizational strategies. For example, the correlations between the observed and reported incidence of the simultaneous instruction of all pupils during second grade instructional activities was .57 for reading and .47 for mathematics (both significant at the .01 level).

An interesting aspect of this analysis was that during reading and mathematics activities at both grade levels adults other than the teacher were observed working in classrooms where the teacher reported that children were grouped instructionally for specific skill need. Only about one-third of the second-grade teachers and about fifteen percent of the fifth-grade teachers reported that they did not teach reading and mathematics to all of their pupils themselves. Therefore, we are discussing a relatively small proportion of the teachers. Teachers were not observed (to a degree where significant correlations emerged) teaching groups of pupils in classrooms where they reported grouping by skill level. It appears that teachers who use other adults as instructors frequently assign the major responsibility for specific skill instruction to the other teaching adults.

The second test of the validity of teacher reports was a comparison of the reported time spent teaching specific reading and mathematics skills compared with observed evidence of pupil demonstrations of those skills.

---

[18] APPLE Observation Variables and Their Relation to Reading and Mathematics Achievement, Nadine M. Lambert and Carolyn S. Hartsough, Volume III, Chapter 1, Phase II Final Report.

The correlations were small and the levels of significance approached the liberal .10 level arbitrarily set by the authors as the cut off point for these analyses. Nevertheless, the reports of teachers about time spent teaching specific skills appeared to be reflected in evidence of pupil skills in those areas.

Teacher reports of their instructional activities, procedures and materials have demonstrated their accuracy and validity in this study. The Phase II teachers who expended so much effort in providing that information should be gratified to know that the results justified their efforts. "Self-reporting" as a reputable form of gathering research information may have gained new lustre as a result of their careful efforts.

The Relationship of Teacher Aptitudes to Teaching Behavior

As part of the investigation of the relationship between teacher characteristics and pupil learning, a study was made to determine the impact of selected teacher variables, such as aptitude, knowledge, attitude, and cognitive style on classroom teaching behavior.[19] Based in part on the findings of earlier research, we hypothesized that instructional behaviors would be influenced directly by teacher aptitudes and that the aptitudes would also affect teaching behavior indirectly through differences in teacher knowledge. Attitudes and cognitive style were also expected to have a direct effect on teaching behavior.

The subjects were the 95 teachers in the Phase II study. All of the teachers took a battery of tests which included measures of aptitude, knowledge, cognitive style and attitude.

---

[19] The Relationship of Teacher Aptitudes to Teaching Behavior, Ruth B. Ekstrom, Volume V, Chapter 1, Phase II Final Report.

The aptitude battery consisted of 15 unifactor tests measuring the cognitive factors known as verbal comprehension; general, inductive, and logical reasoning; associative and meaningful memory; associational, expressional, and ideational fluency; semantic originality; semantic redefinition; sensitivity to problems; and spontaneous semantic flexibility. We also included in this battery the teacher verbal ability test from the Coleman study.[20]

The complexity of the experimental design made it desirable to reduce the number of scores obtained from the aptitude test battery and the other teacher tests. The final derived scores and their composition are indicated in Table 14. Obtaining the four aptitude factor scores presented an interesting problem.

While there was no reason to expect different aptitude patterns at the two grade levels, it was decided that the analyses for the two groups of teachers should be kept separate until it could be demonstrated that these aptitude factors were essentially congruent. At both grade levels there were four aptitude factors with roots greater than one and, at each grade, these factors accounted for 63 percent of the variance. However, the factors were distinctly different in structure regardless of the methodology used to derive or rotate the factors. While it would have been interesting to continue exploring the relationships using these different aptitude factors, we were more concerned at this stage of our research in exploring the similarities in the aptitude-performance relations across the two grade levels. Consequently, a target matrix

---

[20] Equality of educational opportunity, Coleman, J. S., et al., Washington, D. C.: Government Printing Office, 1966.

Table 14

Teacher Test Scores

I. Aptitude Scores

A. Verbal Fluency - A composite score, derived from a factor having its major loadings on:

1) Verbal Facility. The verbal ability test used in the Coleman study. The respondent is asked to select, from among five options, the best word to be used to complete a sentence.

2) Vocabulary (V-4). The respondent is asked to select, from five options, the best synonym for a stimulus word.

3) Topics (FI-1). An ideational fluency test. The respondent is asked to write as many ideas as possible about a given topic.

4) Controlled Associations (FA-1). An associational fluency test. The respondent is asked to write as many synonyms as possible for each of several stimulus words.

5) Making Sentences (FE-1). An expressional fluency test. The respondent is asked to write sentences of a designated length when the first letter of most of the words is specified.

B. Memory - A score derived from a doublet factor with loadings on:

1) Picture-Number (MA-1). An associative memory test which asks the respondent to recall the two-digit numbers paired with each of several pictures previously studied.

2) Sentence Completion. A meaningful memory test in which the respondent is asked to recall the one word which has been deleted from each of several previously studied sentences.

C. Reasoning - A composite score derived from a factor having its major loadings on:

1) Picture Grouping (I-3). A figure classification test. The respondent is asked to determine the rule or reason which determines the assignment of a simple figure to one of two or three groups and then to indicate to which group additional figures should be assigned.

2) Nonsense Sayings (RL-1). A logical reasoning test. The respondent is asked to indicate if the conclusion drawn from two preceding statements shows good or poor reasoning.

3) Inductive Reasoning (I-1). The respondent is asked to determine the rule which relates four groups of four letters and to mark the fifth and unrelated group.

4) Mathematics Aptitude (RG-2). The respondent is asked to select from five options the correct answer to simple word problems which stress reasoning and include some simple algebra.

D. Flexibility - A composite score based on a factor with its main loadings on:

1) Finding Useful Parts (RE-1). A test of redefinition ability. The respondent is asked to select, from five options, the one object which could be used as a makeshift substitute for a specified purpose when the object usually used is unavailable.

2) Story Surprises (O-1). A test of semantic originality. The respondent is asked to write two different and surprising endings for each of several short stories.

3) Listing Objects (XS-3). A test of semantic flexibility. The respondent is asked to list as many things as possible which might be found in a specified setting.

4) Planning Test (Sep-1). A test of sensitivity to problems. The respondent is asked to indicate what is wrong with each of several plans presented for solving a variety of practical problems.

II. Cognitive Style

Score on the Group Embedded Figures Test (Consulting Psychologists Press, Palo Alto, California).

III. Knowledge Scores

A. Teaching Methods - Scores derived from short tests of methods of teaching either reading or mathematics at the elementary school level. Some items include general theoretical background in the subject.

B. Subject Matter - Scores derived from teacher performance on tests of inferential reading and decoding and of simple arithmatic computation.

IV. Attitude Scores

A. Aspirations - A composite score derived from 6 items relating to desire for leadership, recognition, and opportunities.

B. Satisfaction - A composite score derived from 10 items dealing with satisfaction with various aspects of the school, with teaching as an occupation, and with contacts with teachers and admininstration.

C. Perception of Student Characteristics - A composite score derived from 3 items about student educational background, socioeconomic level, and difficulty in controlling students.

based on the hypothesized structure of the aptitude test battery was prepared and the obtained solutions were rotated to this target. The resulting factors, indicated as the aptitude scores in Table 14, all had coefficients of congruence over .90 with the target.

The teaching behaviors (Table 15) were derived from three different sources of data on teachers' classroom behavior. Two of these are direct observational techniques, APPLE and RAMOS. The third method was a Work Diary described in the previous section of this report. These three sources provided 136 scores for each teacher. These were reduced by grouping to 22 scores which can, in turn, be considered to belong to six distinct categories.

The four teacher aptitude scores, the cognitive style score, the two knowledge scores, and the three attitude scores were entered into a path analysis to determine their effect on the teaching behaviors. These data are presented in Tables 16 and 17. The path coefficients can be thought of as being comparable to partial correlation coefficients.

Results. As can be seen from these tables, although many of the paths from the teacher scores to instructional behaviors were large and at or near significance, very few of the teacher scores showed a consistent relationship across instructional behaviors at both grade levels and for both reading and mathematics instruction.

The two teacher scores which did show consistent relationships with instructional behavior were cognitive style, which was negatively related to the social control and management of pupil behavior, and aspirations, which was negatively related to having pupils working independently of the teacher. Thus, the field-dependent teachers were more concerned with

Table 15

Teaching Behaviors

I. Irstructional Time
    Work Diary 1 - An index of the amount of time spent and teaching reading or mathematics
    RAMOS 1 - An index of the variety of instructional roles (high score implies more time in instructional and facilitating roles)

II. Instructional Content
    Work Diary 2 - Range of content or skills taught
    Work Diary 3 - Quality of teacher's conceptualization of instructional activities
    RAMOS 2 - Variety of instructional aims or purposes (high score implies more time in practice and review)

III. Instructional Materials
    Work Diary 5 - Number of different kinds of materials used
    RAMOS 3 - Vareity of materials used

IV. Instructional Organization
    Work Diary 4 - Variety of instructional contexts in the classroom
    APPLE 5 - Pupils work independently of teacher
    APPLE 6 - Teacher directs pupils working in groups
    APPLE 7 - Teacher works with whole class

V. Instructional Activity
    RAMOS 4 - Time spent in instructional activities (seatwork)
    APPLE 8 - Organizing and maintaining the instructional process
    APPLE 9 - Short term interactions (monitoring the progress of pupils)
    APPLE 10 - Sustained interactions to enhance pupil understanding
    APPLE 11 - Location of teacher in the classroom (high score implies greater mobility)

VI. Social Control
    APPLE 12 - Instructional responsiveness to individual pupils
    APPLE 13 - Social control or management of pupil behavior
    APPLE 14 - Nonresponsiveness to individual pupils

Table 16

PATH COEFFICIENTS FROM SECOND GRADE TEACHERS' APTITUDES, COGNITIVE STYLE, KNOWLEDGE, AND ATTITUDES TO THEIR CLASSROOM BEHAVIORS IN TEACHING READING AND/OR MATHEMATICS

Instructional Behaviors

	Time		Content			Materials		Organization				Activity					Interaction		
	WD1	R1	WD2	WD3	R2	WD5	R3	WD4	AP5	AP6	AP7	R4	AP8	AP9	AP10	AP11	AP12	AP13	AP14
READING (n=40)																			
Aptitudes																			
Verbal Fluency	--	--	.27	.33	--	--	.30	-.55	.36	.48	--	--	-.25	--	--	-.28	.49	.25	-.37
Memory	--	--	--	.57	--	--	.53	-.28	.27	--	--	.39	--	--	--	-.37	.33	--	.32
Reasoning	--	--	--	--	.27	-.77	--	--	--	-.29	--	--	--	.66	--	.26	--	--	--
Flexibility	.25	--	.38	-.43	--	.77	-.46	.52	-.92	-.49	.34	--	-.48	--	--	.34	-.25	.30	--
Cognitive Style	--	-.41	-.44	--	-.37	-.37	-.54	--	--	--	--	-.28	.25	-.60	-.46	--	-.42	-.32	--
Knowledge of:																			
Teaching Methods	.36	--	--	-.29	--	-.36	--	--	--	-.26	--	--	--	-.29	--	.44	--	--	--
Subject Matter	-.27	--	--	-.53	--	--	--	--	--	.30	-.42	--	.39	-.45	--	--	--	--	--
Attitudes																			
Aspirations	--	--	--	-.45	--	--	--	.60	-.45	--	.49	--	--	--	--	.31	--	.33	--
Satisfaction	--	-.31	-.58	.46	--	--	--	--	-.28	.36	--	--	--	--	--	--	.38	--	--
Perception of Student Characteristics	-.38	--	--	.31	--	.53	-.29	--	--	--	-.40	--	--	--	--	-.78	--	--	--
MATHEMATICS (n=41)																			
Aptitudes																			
Verbal Fluency	--	--	--	--	--	--	--	-.26	--	.32	--	--	--	--	-.33	--	--	.71	--
Memory	--	--	--	--	-.26	-.46	--	-.50	--	.51	-.31	.35	--	--	.51	.25	.46	--	-.25
Reasoning	--	.56	.42	-.99	--	--	.46	.41	--	-.27	--	--	--	.32	--	--	--	--	--
Flexibility	.59	-.30	.36	--	--	.40	-.40	.99	-.65	-.68	.86	--	-.36	--	--	--	-.36	--	--
Cognitive Style	--	-.37	--	.25	-.33	--	-.31	-.59	--	--	--	--	--	-.34	--	--	--	-.51	--
Knowledge of;																			
Teaching Methods	--	--	--	--	--	-.30	--	--	--	.79	-.61	--	--	--	--	--	.29	.27	--
Subject Matter	--	-.32	-.50	.38	--	.32	--	--	.27	--	--	--	--	-.57	--	--	-.58	--	.37
Attitudes																			
Aspirations	.33	--	--	--	--	.57	--	.67	-.36	-.54	.51	--	--	--	--	--	-.30	--	--
Satisfaction	--	--	-.30	--	--	--	--	-.93	--	.46	--	.28	--	--	--	--	--	--	-.45
Perception of Student Characteristics	-.28	--	--	.67	--	--	--	-.27	.44	--	--	--	.28	--	--	--	.31	--	.35

(Coefficients of .25 or greater)

Table 17

PATH COEFFICIENTS FROM FIFTH GRADE TEACHERS' APTITUDES, COGNITIVE STYLE, KNOWLEDGE, AND ATTITUDES TO THEIR CLASSROOM BEHAVIORS IN TEACHING READING AND/OR MATHEMATICS

Instructional Behaviors

	Time		Content			Materials		Organization				Activity					Interaction		
	WD1	R1	WD2	WD3	R2	WD5	R3	WD4	AP5	AP6	AP7	R4	AP8	AP9	AP10	AP11	AP12	AP13	AP14
READING (n=53)																			
Aptitudes																			
Verbal Fluency	--	-.25	.47	.71	-.44	-.73	--	--	.41	-.29	.25	--	.63	.31	--	-.43	--	--	--
Memory	-.33	--	-.26	--	--	--	--	-.34	--	--	--	--	--	--	--	--	--	--	--
Reasoning	-.29	--	-.30	-.40	--	--	--	-.54	.30	--	--	--	--	-.39	--	.28	--	.49	--
Flexibility	--	--	--	-.58	--	--	--	.31	--	--	-.37	--	-.50	--	-.29	--	-.52	--	--
Cognitive Style	-.33	--	.26	--	.45	--	--	--	-.37	--	.31	--	--	.25	.45	--	.51	-.25	--
Knowledge of:																			
Teaching Methods	.38	--	--	--	--	--	--	.34	--	.31	--	--	.28	--	.44	-.35	.30	-.31	--
Subject Matter	--	--	--	.25	--	.72	--	--	-.38	.33	--	--	-.27	--	--	.37	--	--	-.30
Attitudes																			
Aspirations	--	--	.32	--	--	--	--	.48	-.36	--	--	--	--	.40	.36	-.48	--	--	--
Satisfaction	-.35	--	--	-.29	.40	--	.38	-.32	--	-.46	--	--	--	--	-.27	.59	.29	--	.28
Perception of Student Characteristics	--	--	--	--	-.37	--	-.44	--	--	-.39	--	--	-.36	--	--	--	--	.32	--
MATHEMATICS (n=54)																			
Aptitudes																			
Verbal Fluency	--	--	.53	--	--	-.32	--	.45	--	--	--	--	--	-.25	.28	--	--	-.36	--
Memory	--	--	--	--	--	--	--	--	.28	--	--	-.26	--	.42	--	--	.34	-.29	.28
Reasoning	-.35	--	-.36	-.51	--	--	--	--	--	--	--	--	--	--	--	.47	--	.26	--
Flexibility	--	--	--	-.38	--	.29	--	--	--	--	--	--	--	-.39	-.25	--	--	.71	-.40
Cognitive Style	--	--	--	.42	--	--	--	--	--	--	--	.34	-.33	--	.60	-.25	--	-.33	.31
Knowledge of:																			
Teaching Methods	--	-.27	--	.52	--	--	--	--	--	--	--	--	--	.43	-.33	--	--	--	--
Subject Matter	--	--	--	.28	--	--	--	--	--	--	--	--	--	.40	--	--	.27	--	--
Attitudes																			
Aspirations	.27	--	--	--	-.27	--	--	--	-.25	--	.32	--	--	--	--	-.37	--	--	--
Satisfaction	--	--	--	--	--	--	.26	--	--	-.30	--	--	--	.39	--	--	.33	--	.33
Perception of Student Characteristics	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	-.30	--	.43	--

(Coefficients .25 or greater)

maintaining behavioral control in the classroom than were the field independent teachers. Teachers with low aspirations were more likely to use a classroom structure in which the pupils worked independently of the teacher.

With these two exceptions, the analysis did not show any other teacher score which was consistently related to a teaching behavior across both grade levels and both subjects. However, there is a strong suggestion of interaction between specific teacher characteristics, especially the aptitude scores, and instructional behavior. These data suggest that teachers perceive the demands of these grades and subjects differentially and do not regard teaching as a homogeneous task. Thus, it can be hypothesized that teachers may select different teaching styles according to their perceptions of the demands of the instructional task.

One of the most dramatic of the aptitude-behavior interactions is that between the flexibility factor and instructional organization at the second-grade level. The more flexible teachers utilized a more complex classroom organization (WD-4), which implies they were more likely to have sole responsibility for instruction. These more flexible teachers tend to teach the class-as-a-whole (AP-7), do less independent pupil work (AP-5) and do less instruction in groups (AP-6). This result is interpreted to indicate that more flexible teachers are better able to respond differentially to pupils without having to resort to using various organizational strategies (aides, groups, etc.) to produce this individualization.

Verbal fluency shows a relationship to the quality of teaching methodology while reasoning ability, as defined by these measures, seems to be negatively related to the quality of teaching behavior.

These data suggest that further research exploring the predictability of instructional behaviors from teacher aptitudes, knowledge, cognitive style, and attitudes would be promising.

Some Determinants and Consequences of Teacher Expectations Concerning Pupil Performance[21]

In the near decade since the publication of Pygmalion in the Classroom, the research literature has been flooded with sources of studies attempting to replicate or to explain the seminal Rosenthal and Jacobson finding. Several recent reviews of these studies have reiterated the fairly consistent finding that experimentally produced teacher expectations, or "bias," do effect both the way in which the teacher interacts with the pupil and the pupil's ultimate performance. Furthermore, research shows that self-induced teacher expectations, or "expectancies," are correlated with both teacher behavior and pupil performance.

What is conspicuously absent in this literature, however, are studies addressed to two underlying problems: (1) what are the determinants of self-induced teacher expectations, and (2) what is the magnitude of the consequences of such expectations on pupil performance? It will be the purpose of this section to investigate these issues using the framework of a longitudinal, non-experimental design, and a production function analysis.

Research on teacher expectations typically has addressed one of the three following questions:

---

[21] Teacher Expectations and Pupil Performance, Marlaine Lockheed, Volume V, Chapter 2, Phase II Final Report.

1. Do variations in teacher expectations provoke differences in pupil performance?
2. How are teacher expectations communicated to students?
3. Do various student characteristics, such as race, sex, or social class, provoke differences in teacher expectations?

Past research evidence (reviewed in Volume V, Chapter 2) shows that certain student characteristics determine teacher expectations and teacher behaviors, that teacher expectations determine student academic achievement and teacher behaviors, that student characteristics are correlated with academic achievement, and that teacher behaviors affect student academic achievement. In addition, studies of the determinants of self-evaluations show that status characteristics, the evaluations of others, and achievement affect self-evaluations. Self-evaluations have been shown to affect achievement. While the above linkages underlie the conceptualization of much of the reported studies on teacher expectations effects, previous studies have typically explored only one or two of the linkages at the same time.

While the present study will not be able to examine the teacher behavior part of the preceding model (because the available teacher behavior data has been aggregated by classroom rather than by target pupil), the remainder of the linkages will be thoroughly examined.

Method. In this study, students for whom complete data sets were available for both fall and spring data were selected for analysis. A multiple regression technique was utilized, with the teacher expectation for the student's achievement as one dependent variable and with the student's spring achievement as a second dependent variable. Both raw spring achievement and residual gain scores were examined; in this case, residual gain score was the computed difference between the student's actual spring score and the spring score that would be predicted for the student based upon the fall achievement score.

In all cases, analyses were conducted separately by grade and by mathematics or reading subject matter. Thus, without discussing alternate predictor variable combinations, this section summarizes a minimum of 12 regression equations.

Teacher expectations were measured by asking the teacher to rank order each child in the class according to how well she or he thought the child would do in reading and in mathematics during the school year. Ranks were adjusted for class size by converting them to deciles.

Student characteristics which were thought to possibly influence teacher expectations included the child's sex, racial identity, estimated socioeconomic background, physical handicaps, participation in compensatory education programs and the child's ability to speak and understand English.

Results. The first question to be posed was: "What individual pupil characteristics determine the teacher's expectations for each child?" To investigate this question, a series of stage-wise multiple regressions were run. Included as independent variables in these equations were the student's achievment, the demographic variables mentioned above, the student's self-expectations, attitudes toward the subject matter and the expectations of his or her peers. Various combinations of these variables were entered into the regression. This discussion will describe the results of only one of these combinations: all variables except student achievement entered on the first stage, and then student achievement entered on the second stage. The results show that demographic variables alone account for little variance in teacher expectations; the variance explained lies between 13 and 16 percent of the total variance.

When the effects of student expectations, attitudes and achievement were considered simultaneously with demographic or status variables, student achievement was the strongest single predictor of teacher expectations. After the effect of student achievement was partialled out of the equation, and the remaining variables were considered simultaneously, it was found that:

1. The sex of the student was a significant correlate of teacher expectation for reading at both grade levels, but not for mathematics, after other variables were partialled out;

2. Being Black was not a correlate of teacher expectations in second grade, while it was a correlate in fifth grade. In the fifth grade, teachers held higher expectations for Black students than for white students;

3. Being Asian-American was positively correlated with teacher expectations;

4. Being Mexican-American was positively correlated with teacher expectations for second-grade mathematics and fifth-grade reading;

5. Having a physical handicap was negatively associated with teacher expectations for second grade but not for fifth grade;

6. In general, having participated in compensatory education programs was negatively associated with teacher expectations. The only exception is having participated in a Miller-Unruh reading program in second grade;

7. Social class was not related to teacher expectations;

8. Teachers held higher expectations for bilingual students; and

9. Students' own expectations and attitudes were positively correlated with teacher expectations.

The second question posed was: What is the effect of teacher expectations on change in student achievement? To investigate this question, both spring scores and predicted spring scores were used as dependent measures, and the effects of teacher expectations were examined for the total population under consideration. In addition, a separate group of students who had scored within one standard deviation of the mean on their fall achievement were separately examined. Results of this analysis show that:

1. Teacher expectations were modestly but consistently and significantly related to spring student achievement when the effects of entry level skills had been partialled out. Teacher expectations accounted for from three to nine percent of the variance in spring achievement, and that relationship was consistent for both second and fifth-grade students in both reading and mathematics;

2. Although small, the contributions to the variance in student academic growth provided by teacher expectations were greater than the contributions of the more conventional demographic indices such as sex, race and social class;

3. When the effects of teacher expectation on residual gain scores were examined, three to nine percent of the variance in gain was explained by teacher expectations; and,

4. High teacher expectations were estimated to increase the achievement of the average student as much as one standard deviation, when compared to similar students for whom teachers held low expectations.

Summary. This study has addressed the teacher expectation problem using methods which differ from those used in previous research. Previous research has focused on changing teacher expectations experimentally and observing differences in student outcomes; teacher expectations were not manipulated in this study. In studies in which teacher expectations were measured rather than manipulated, student performance outcomes have not been measured or controlled for; in this study, teacher expectations were measured and student performance outcomes observed at two points in time.

The findings suggest that self-induced teacher expectations are generally well-founded, and are not negatively biased agains minority students, or against males or females. Also when teachers hold higher versus lower expectations for similar average students, the differences in their subsequent achievement levels can be as much as one standard deviation apart.

These findings suggest that although the correlations between teacher expectations and subsequent student achievement are low, they should not be overlooked, as the impact on student achievement can be substantial.

PHASE II SPECIAL RESEARCH STUDIES

The Evaluation of Videotape Recordings of Teacher Behavior[22]

Teaching behavior was investigated through the observation of videotape recordings. Project videotapers recorded the teaching of the teachers taking part in the BTES study. Most were taped three times on one visit, but a subsample of 12 teachers was taped on three or four separate occasions. The purpose of the subsample was to provide information about the consistency of observed behavior over time.

---

[22] The Evaluation of Teaching Behavior Through Observation of Videotape Recordings, Jonathan Sandoval, Volume III, Chapter 3, Phase II Final Report.

Several instruments were created or adapted for use in analyzing the videotape recordings. Two new lexicons for observation were developed, the Teacher Instruction Lexicon and the Teacher Feedback Lexicon, which are devoted primarily to the observation of teaching behaviors. The Teacher Instruction Lexicon is a listing of the various categories of teacher behaviors used in conducting a lesson to one or more children in the classroom: introductory remarks; explanations or expository remarks a teacher might make in the course of presenting information to be learned; and, questioning or testing behaviors a teacher might use to stimulate mastery learning in pupils or to evaluate what learning has taken place.

The Teacher Feedback Lexicon is an exhaustive listing of the ways in which teachers report information back to pupils about the pupil's performance in learning or in following classroom rules. Across the three general categories of (1) Positive Responses, (2) Non-Committal Responses and (3) Negative Responses, five kinds of responses were recorded: simple informational responses, responses that are "Socratic" reactions, responses that are praise or punishment acts, responses that are attempts to create peer involvement, and responses that involve ignoring.

Two Videotape Rating Scales were also developed. The items on these schedules are concerned with the pace, tempo and variety in the lesson, the teacher's movements and enthusiasm, the organization and emphasis of the lesson, the teacher's empathy for the students, the pupils' reaction to the lesson, the teacher's questioning strategy, and the Rogerian concepts of warmth and honesty. Observers rated performance after coding each reading and mathematics tape with the Teacher Feedback Lexicon.

Because the lexicons were new and needed to be validated, a standard observation system was employed, Flander's Interaction Analysis, which could be used to serve as a benchmark against which to evaluate the utility of the new systems. All of the instruments were used in coding the videotapes recorded in the subsample classrooms. Only the new observational or rating scale items that had demonstrated inter-rater reliability were used to evaluate classroom consistency. In general, most of the observed and rated teaching performances were stable over time. The rating and observation systems were then evaluated and simplified for coding tapes from other project classrooms.

Results. What kind of information can be obtained by observers viewing the videotapes? Some examples from the descriptive sections of the report will be used to illustrate the different kinds of teaching behaviors observed.

How common are various instructional behaviors? Teachers at the elementary level spend most of their time giving facts, asking simple questions and giving directions about how to do something. On the other hand, the teachers rarely stimulate recall of prerequisite skills in the process of giving a lesson, rarely give application of ideas they are teaching, rarely contrast and compare ideas, almost never ask for negative opinions, and do not often attempt to determine if a child has mastered a concept.

Reading and mathematics lessons appear to differ in the instructional events observed. There are more explanations given in mathematics lessons whereas in reading there is more cueing and asking follow-up questions for additional information. There does not seem to be much difference in the frequency of the various instructional events in the second and fifth grades. The substance, the content of the events, undoubtedly does differ between grade levels, but not the absolute occurrence.

Teachers in lower class schools seem to differ from other teachers in that they are less prone to state facts, are less likely to give explanations and are less likely to ask if information is understood. In general there seems to be somewhat less verbal behavior in these classrooms than in others. Teachers in upper middle class schools, on the other hand, state facts more often, ask more frequently if information is understood, and engage in less direct cueing than do other teachers.

What feedback techniques are commonly used by teachers? The most common reaction to positive achievement in the classroom is to inform the pupil simply that he is right, without much embellishment. General praise is a distant second in frequency as a response to positive behavior.

When there are both right and wrong elements in a child's answer to a problem or a question, it is common for the teacher to attend to the negative aspect of the response or to simply acknowledge hearing the response. In giving feedback teachers also, but less frequently, attend to only the positive aspects of the response and often ask the child to elaborate.

Reacting to poor acheivement, that is to a poor answer or an incorrect answer to a question, teachers most often ask a child to try again, tell him he is wrong, or supply him with the right answer. Most puzzling is the fact that there is very little guidance given when errors are made. Second-grade teachers are more likely to ignore positive achievement and to direct a child to try again, should he fail. Fifth-grade teachers, however, tend simply to acknowledge hearing a mixed positive and negative pupil response. Fifth grade teachers are more likely to reprimand a pupil for an academic failure than are second-grade teachers.

It appears that teachers of lower class pupils seem to give more simple knowledge of results and fewer elaborated knowledge of results for achievement. They also use more general praise than do other teachers. They ignore positive achievement less, and they more often ask pupils to elaborate when there are both positive and negative aspects to what they have done. Some teachers in upper-middle class schools are more likely to reprimand a child for failing than are other teachers, which may reflect high expectations for performance.

Although teachers do tell pupils that they are correct when they are behaving well, they are much more likely to react or attend to negative behavior. Some teachers do attend to the positive when both positive and negative behavior occurs. The usual reaction is a verbal reprimand, directions to stop the behavior. Very few teachers purposefully ignore negative behavior, and very seldom does a teacher involve peers in giving feedback to children whose behavior was inappropriate.

Again there are few differences between second and fifth-grade classrooms, the only notable one being the fact that misbehaving second-grade pupils are told more often to proceed with an assigned task than are fifth-grade children. Misbehavior is reacted to in the same way, whether in a reading lesson or in a mathematics lesson.

There are more feedback events of all types observed in the tapes of teachers in lower class schools, reflecting the possibility that there is more nonachievement oriented behavior in these classrooms. There is more praise and recognition for good behavior in the lower socioeconomic status (SES) classrooms, but there is also more attention to negative behavior and more directions to get on with an assigned task.

The information contained in the preceding paragraphs needs to be evaluated in future research. The fact that a particular event is infrequently observed does not mean that the event is not an important behavior. In many cases, the very fact that a teacher exhibits a particular behavior is more important than the number of times he or she performs that behavior. The frequency data, then, must be used with great care in designing future experimental research.

Do teaching behaviors relate to pupil achievement and attitude? The relationship between the teacher behaviors obtained from the videotapes and pupil achievement was estimated by correlating the observational variables with the mean adjusted spring achievement and attitude measures across the total group of second and fifth-grade classrooms.

It should be noted that we are correlating variables derived from the viewing of one 40 minute segment of videotape collected at one time during the school year with measures of pupil growth across an academic year. To expect many meaningful relationships to emerge from such a limited sample of a teacher's behavior is extremely optimistic. Any relationships which do emerge should be considered very seriously as important aspects of teacher behavior to be further examined or explained. For this partial correlation analysis, the Teacher Instruction observational variables collected during reading lessons were related to the California Achievement Test (CAT)-Reading Comprehension and to the ETS developed Reading Attitude measure. The observational variables obtained from the mathematics lesson tapes were related to the CAT-Mathematics Concepts, the CAT-Mathematics Computation, and the ETS developed Mathematics Attitude measure. Three of the Teacher Instruction Behavior categories were related to achievement or attitude toward subject matter.

During fifth-grade reading instruction, the more expository remarks made by the teacher, the better is the pupil's attitude toward reading. Evidently the more the teacher talks about the reading assignment and gives facts, examples, summaries, and applications, etc., the better the pupils feel about reading and reading lessons. On the other hand in mathematics instruction, at the second grade, the more a teacher makes expository remarks, the less well the pupils do at acquiring mathematical concepts. Perhaps listening to the teacher and working with materials may be mutually exclusive in a second-grade classroom.

Giving Directions is negatively correlated with both achievement and attitude almost across the board at the fifth-grade level. The more directions a teacher gives, the less well the class, as a group, performs. A correlation coefficient does not imply causality; we might also hypothesize that the less well the class does, the more directions may be given because the pupils are not working at the correct level or are not able to do the work on their own.

Asking General Questions, too, is negatively related to mathematics achievement and attitude toward mathematics at the fifth grade. The more verbal behavior in the form of questioning (and in mathematics this would probably be in the form of asking children to recite answers) at the fifth-grade level, the poorer the pupils' achievement and the more negative the pupils' attitude toward mathematics. If drill and poor achievement are related, one would hypothesize that the more covert the practice and the more individualized the instruction, the better the achievement. This result should be explored further.

Asking Questions is positively related to gains in reading comprehension at the second-grade level, however. The technique of questioning children about what they have read is a standard one for stimulating discussion which fosters comprehension. More research should be done to determine the nature and timing of questions which are most conducive to the development of reading comprehension skills.

Only one of the 12 Teacher Feedback event categories variables, Calling Attention to Error, is unrelated to any of the pupil growth measures. Simple Knowledge of Results during reading lessons is negatively related to reading comprehension at the fifth grade. The more pupils are told they are correct when they are, the less well they do on the measure of reading comprehension. Perhaps pupils improve in comprehension more in the fifth grade if they are given questions which do not involve right and wrong answers, but which tap pupil opinion and interpretation of reading material.

Asking a Child to Elaborate after he has responded seems to be related to reading comprehension at the second-grade level. This technique of having a child talk more about his thoughts is evidently an effective one for raising achievement in reading comprehension.

Another teaching technique positively related to achievement in the second grade is using other pupils to give praise, encouragement and knowledge of results to children. Encouraging Other Pupils to Give Feedback is positively related to both reading and mathematics achievement, as well as attitude toward mathematics. At the fifth grade, however, having other pupils give feedback is negatively related to reading achievement. Having others respond to your answers or products during a reading lesson may be embarrassing and create a climate unconducive to learning at this later age.

Ignoring Positive or Mixed Achievement may be thought of in two ways. A teacher who ignores a pupil's achievement may be communicating a lack of interest in the product, creating an unpleasant feeling in a pupil, or may be communicating a set of high standards and expectations which would stimulate a pupil onward. At the fifth grade, this variable is negatively related to attitude toward reading. Perhaps fifth graders interpret ignoring as disinterest. In the second grade ignoring was positively correlated with mathematics achievement. At the second grade, ignoring may be related to high expectations for the pupil. These explanations are worthy of testing in future investigations.

The variable, Weak Acknowledgement of Achievement, is a measure of how much a teacher reacts to a pupil achievement that is neither completely right nor wrong. Reacting to a mixed pupil response with some acknowledgement is positively correlated with achievement in both reading and mathematics for second graders. The younger pupils seem to profit from having a reactive, responsive teacher. This variable is negatively correlated with mathematics attitude in fifth-grade pupils, however.

Encouraging a Pupil to Try Again is an important teacher behavior; if the pupil does not attempt a task again until he gets it correct, he will not learn. Trying Again is negatively related to mathematics achievement for second graders but is positively correlated with attitude. This paradoxical finding should be validated in future research.

How do teachers' reactions to non-academically oriented behavior relate to achievement and pupil attitude? Praise for good behavior and Giving Warning for negative behavior has little or no effect on fifth-grade pupils' achievement or attitude. Praise seems to have a positive effect on second graders' reading attitude and mathematics computation skills. The more punishment the teacher gives, however, the poorer the pupils are apt to do in mathematics. It seems reasonable that the more disruptive the mathematics class, the less a pupil will learn of mathematics concepts. However, the more Warning, the more a teacher has only to remind children of consequences rather than punish them, the more mathematics learning takes place.

It is important to note that none of the event categories are significantly related to achievement or attitudes across both grade levels. Few of the event categories are related in the same way to both achievement and attitudes in a subject, and few are related in the same way to mathematics and reading. Teacher feedback behavior does not have universal effects. Rather, certain types of feedback have effects in certain situations and not in others. Any further research on feedback must take into account more carefully the context and appropriateness of the feedback behavior.

In the fifth grade, for example, none of the correlations between feedback behavior and the criterion variables are positive. The more feedback pupils are receiving in the fifth grade, the worse is their achievement and attitude. While feedback may assist second-grade pupils, the presence of feedback in fifth-grade classrooms seems either to alienate pupils or to be an indication that the pupils are not performing well. Pupils who are functioning well apparently do not need feedback at the fifth grade. When children are being rewarded intrinsically by the materials and activities they are involved with, they are learning. When pupils are not achieving well, teachers may resort to more verbal, overt feedback attempting to influence learning.

All of the rating scale items are related to pupil achievement or attitude except Teacher Gestures. Teacher Movement is correlated with positive pupil attitude toward reading in second-grade classrooms. To receive a high rating on this aspect of teacher variability, teachers usually move about the room, interacting with instructional materials, for example, by helping students act out stories.

Teacher Enthusiasm, ordinarily a powerful variable in research on teaching, is related only to attitude toward mathematics at the second grade. In this case, the more enthusiastic the teacher, the less favorable is the attitude toward mathematics.

The rating of Time Spent on Learning Activities is positively correlated with second-grade attitude toward reading and mathematics achievement. The more time spent in learning in the second grade the better the children feel about reading and the more mathematics concepts they learn. Clearly attitude and achievement are related. When teachers give sufficient time for pupils to master materials and to succeed with them, the pupils learn and enjoy learning. A similar explanation may serve for the obtained relationship between fifth-grade reading achievement and attitude and the rating of Lesson Learning Centered. Apparently fifth graders need task-orientation both to learn and to feel good about reading; when fifth-grade reading lessons are focused on reading, progress occurs.

The rating of the extent to which Learning Activities are Differentiated for pupils, i.e., the amount of observed individualized instruction, is positively related to second-grade mathematics attitude. The more teachers individualize instruction in mathematics at the second grade, the more positively the pupils feel about the subject. Fifth-grade pupils, however, seem to have a more negative attitude toward reading if there are differentiated reading assignments in their classrooms. At the fifth grade, pupils are much more aware of peer attitudes and those who find themselves assigned to material below grade level may feel insulted and discouraged about reading.

The technique of Delaying Naming a Child to answer a question after it has been asked seems to be an effective one in increasing reading comprehension at the fifth grade, although it does not create a positive attitude toward reading. At the second grade children are much more eager to be chosen and to volunteer, and are less fearful of failure so that this practice does not have a negative influence on these children.

Observers made judgments of these important global human characteristics displayed by the videotaped teachers, Warmth and Honesty. These are very general and somewhat diffuse traits, but an analysis of these characteristics may suggest important variables. The results of this exploratory study indicate positive correlations between Warmth and achievement in learning mathematics concepts at both grades, between Honesty and achievement in mathematics computation at both grades, and between Honesty and attitude toward mathematics at the fifth grade. Warm teachers do not threaten children and are accepting of children, and this disposition seems to have an effect on learning. Honest teachers give direct feedback, letting pupils know where they stand. Honest teachers do not give gratuitous praise, an attribute which older pupils may appreciate more than younger ones.

One of the goals of this year's phase of the BTES project was to generate hypotheses for further study. All of the significant results obtained and discussed in the report represent specific hypotheses to be tested, to be verified, in more controlled settings. In general, it is quite clear that teacher behaviors which are efficacious at one grade level are not necessarily effective at another grade level, and behaviors related to reading achievement or attitude do not necessarily have anything to do with behaviors related to mathematics achievement or attitude. For example, while Teacher Feedback categories correlated positively with the criteria at the second grade, they correlated negatively at the fifth; while Asking General Questions was a Teacher Instruction behavior that was negatively related to mathematics learning, it was positively related to reading achievement. Because teacher feedback is effective in the second grade and not at the fifth, perhaps social reinforcement from the teacher is much more important to second graders, and reinforcement from success at the activities themselves is more important at the fifth-grade level. One might then predict that the teacher as a "person" is important to second graders, and the teacher as a "planner" or "leader" is important to fifth graders.

Summary. Videotape recording does seem to have a place as a tool for studying teacher behavior. For some purposes, such as looking in detail at certain teacher behavior, the advantages of observation through videotapes do outweigh the disadvantages. No doubt this method will continue to be used in research studies of teacher behavior.

Historical Data on Pupil Performance

Another special project was the study of historical[23] pupil test data for Phase II participant teachers. The purpose of the study was to collect historical data on the reading and mathematics achievement, in previous years, of pupils in the classrooms of the participating teachers, and to compare these historical data with data gathered through the administration of pupil outcome measures during Phase II. The staff collected available standardized achievement test data in reading and mathematics for students who were in the classroom of the BTES teachers during the 1971-72 and 1972-73 school year.

The only case where sufficient test data were available for comparison was in grade two reading achievement for the spring of 1972 and spring of 1973. Twenty-four teachers in five districts were included in the spring 1973-spring 1974 comparisons. Spring 1972 test data were available for ten teachers in three districts. Comparisons were made between spring 1973-spring 1974, spring 1972-spring 1974, and spring 1972-spring 1973 mean scores.

---

[23] Historical Data Study, Patricia Wheeler and Patricia Elias, Volume V, Chapter 3, Phase II Final Report.

For this hypothesis-generating phase, the focus of the stability substudy was to examine existing historical pupil achievement data and to determine whether those data could be used in conjunction with the project data collected in 1973-74 to evaluate the stability of teacher performance over a three-year period. Surprisingly little standardized achievement test data was routinely collected in basic academic areas in the second and fifth-grade classrooms. Post-test or output information appeared to have been collected primarily when it was mandated by the California State Department of Education. It was not mandated at the fifth-grade level. Pre-test or pupil input information was either not available or was inaccessible one or two years after having been collected. Any interpretations about the effectiveness of the teachers and their previous pupils or about the stability of impact of the teachers on pupil outcomes over a period of three years must be made with the understanding that the results are based on pupils of unknown aptitude and entry level skills.

Based on the data on reading achievement at the second-grade level that were analyzed, correlations between the years of historical information about pupil performances suggested that the performance of the teachers' pupils tended to be stable over time. However, when the data were organized in ways that began to reflect something about the characteristics of the pupil populations, the correlations suggested that year-to-year relationships among performances varied substantially by type of geographic location and by SES level of schools. It was not possible to tell from the types of information available whether these variations were due to pupil differences, teaching differences, or both, but it is reasonable to assume that the demographic indices represent pupil population differences to some extent. The analysis of the data utilizing fairly rough demographic indices of school characteristics as a proxy for pupil characteristics illustrated the fallacy in thinking about teaching performance in terms of the single available indicator of pupil outcomes.

A Videotape Test of Teachers' Diagnostic Skills

One of the special studies for Phase II of the Beginning Teacher Evaluation Study was to develop a videotaped test of teachers' skills in diagnosing the reading difficulties of children.[24] The test was a series of episodes in which children were depicted reading. Each child demonstrated some difficulty in performing a particular reading skill. The reading skills included were similar to those identified in the Phase II reading matrix and measured in the Phase II reading test battery for pupils. Second and fifth-grade teachers were asked to diagnose each child's problem and to prescribe an appropriate strategy or strategies for correcting the problem.

Test Development. Two types of content, word lists and paragraphs, were included in the test. Four presentations of each type of content were developed to represent the common types and proportional frequency of errors identified in previous research. The test reflects our bias that elementary school teachers at all levels are challenged by the need to help children at all stages in their reading development, children whose skills range from the pre-primer level to fairly sophisticated levels of reading ability. Therefore, we concentrated on developing a single test which would begin to reflect the range of oral reading problems represented in elementary school classrooms.

---

[24] Reading Difficulties: Videotape Test of Teachers' Diagnostic Skills, Patricia Elias and Gail Hare, Volume V, Chapter 4, Phase II Final Report.

To keep the content of the presentations as logically and psychometrically separate as possible, materials were developed which featured each type of criterion error in single filmed vignettes. Each word list was used to reflect a criterion decoding error; each paragraph was used to reflect a criterion comprehension-related error. Four dominant error types were selected for each of the two kinds of content.

The developmental procedures were similar for the word lists and the paragraphs. The developmental sequence involved, first, the review of the research literature on oral reading errors, noting types of mistakes. It proceeded through categorization of the errors, and focused on those which decline with age or proficiency. The next major step was to select four "major" types of errors and "distractor" or "minor" errors for each type of content (word list or paragraph). The final steps in the developmental sequence were to review related tests and diagnostic materials and to select materials for inclusion in the eight vignettes of the Diagnostic Film Test.

Error Types Selected for Inclusion. After review and categorization, the following types of "major" and "distractor" or "minor" errors were selected for inclusion in the Diagnostic Film Test.

I. "Major" Types of Paragraph Errors

   A. Omission of words, letters
   B. Substitution of graphically similar words that do not make sense
   C. Substitution of phonologically similar words that do not make sense
   D. Substitution of words that are phonologically and graphically similar that do make sense

II. "Minor" Types of Paragraph Errors

   Note: These are less serious errors in the context of reading considered as a progressive learning process in which children approximate words using graphic clues, meaning context and grammatical context.

   A. Hesitation
   B. Repetition
   C. Insertion

III. "Major" Types of Word Decoding Errors

A. Initial word sound errors, primarily consonant related
B. Final word sound errors, primarily consonant related
C. Medial word sound errors, primarily vowel related
D. Medial word sound errors, primarily consonant related

IV. "Minor" Types of Word Decoding Errors

A. Letter sequence reversal errors
B. Single examples of each of two types of "major" errors complementary to the type emphasized in the word list

In each of the word list readings, six examples of the "major" errors are presented, and there are three single occurrence "minor" errors. Four examples of each of the "major" errors occurred in each of the paragraph readings. The three "minor" errors for the paragraph readings occurred once per paragraph. Following each reading of a paragraph, the child was asked four questions, two of which involved literal comprehension and two of two of which involved inferential reasoning. On the videotape, either the two literal questions or the two inferential questions were answered incorrectly by the child reading the paragraph.

Test Format. The printed materials teachers used during the Diagnostic Film Test included three parts: Section One accompanied the films of children reading word lists; Section Two went with the reading of the paragraphs; and, Section Three was presented with a replay of the fourth and fifth taped vignettes. Each written section included a correct version of the materials read in each filmed presentation for teacher use. The response mode built into Sections One and Two was open-ended questions. They were:

1. What is (child's name)'s reading problem or problems? Be as specific as you can.

2. What strategies or approaches would you use to help (child's name) improve his/her reading skills?

3. What resources (human or material) would you seek in your school or district to help you carry out a program for this pupil?

   Materials:

   Human:

4. What resources would you need (if any) that are not presently available in your school or district?

   Materials:

   Human:

The presentations were designed so that when teachers had viewed the eight reading sequences and responded in Sections One and Two of the printed materials, those materials were collected. The videotape was then rewound to present the last of the word lists and the first of the paragraph reading film segments. Teachers were given Section Three and were asked to respond to written multiple choice or recognition format questions about the two presentations. The "major" and the "minor" error categories were included in the multiple choice lists. The multiple choice format was similar to that used in paper-and-pencil tests of teachers' knowledge. The Diagnostic Film Test was designed to more closely approximate the kinds of experiences teachers have while teaching reading than do paper-and-pencil tests of diagnostic skills. We were interested in the differences, if any, between knowledge and application of that knowledge.

Sample. The objective for the Diagnostic Film Test, during this year of instrument development and the generation of hypotheses, was to pilot test the instrument on a sample of Phase II teachers. A sample of 35 teachers was chosen which included 18 at the second and 17 at the fifth-grade level. The Diagnostic Film Test was administered to the selected sample of teachers as part of their spring testing battery.

Scoring the Test. Scoring procedures were developed for each of the three sections of the Diagnostic Film Test. Basic guidelines were established for scoring the test prior to the test administration. Actual Phase II teacher responses were recorded which indicated skills in the diagnosis and remediation of the reading difficulties presented by each of the eight children in the taped vignettes. Their responses were used to illustrate the scoring models for the test.

The scoring standards were fairly rigorous. Teachers were required to identify the major difficulty depicted and to suggest some strategy or strategies for correcting that problem. Credit was also given for the identification of the minor problems depicted and their remediation strategies. Teachers were given credit for any and all strategies they suggested which were associated with a major reading problem. They were credited with only one strategy for each minor error they identified, no matter how many they provided. We did not impose our notion of what a "good" remediation strategy might be.

It was not required that teachers respond in exact "reading specialist" terminology. For example, if a teacher indicated that a child was having problems with the sounds in the middle of words such as t or p, that was accepted as an equivalent response to the identification of the child's problem as medial consonant errors. The teachers occasionally supplied diagnostic information in the remediation sections of the test. Acceptable responses were scored no matter where teachers happened to record them within the test booklet.

Results. The scores of teachers were organized to reflect scores on each part of the test, totals on various combinations, and the total test. One of the areas we wanted to explore was the relationship between teachers' responses to open ended questions and to closed-ended questions in a recognition format. The multiple-choice format is the most frequently used convention in tests of diagnostic skills.

Scores on the open ended decoding questions were not related to scores on the multiple-choice decoding problem for either second or fifth-grade teachers; nor were scores on the open-ended comprehension problems related to scores on the multiple-choice comprehension questions. The multiple-choice decoding problem was not related to the total test scores of either group, but the comprehension questions in that format did correlate with total test scores. This result suggests that the teacher scores on the multiple-choice questions were contributing little or nothing to understanding teacher performance on the test.

In general, the means of the second-grade teachers were higher than those of the fifth-grade teachers. On the free response decoding sections of the test in particular, the second-grade teachers were more able to diagnose accurately and to suggest strategies for remediation of oral reading problems than their fifth-grade colleagues. The scores of the two groups on the open-ended comprehension sections were more similar. The relatively better performance of teachers, in relation to possible points obtainable, on the multiple-choice questions, and the lack of correlational relationships with free responses suggests that performance may be different on the two types of test formats. Further research is warranted to test the effects of test format on teacher demonstrations of diagnostic skills.

The research might include verification of test results in the classroom where, presumably, teaching performance is more dependent on the kind of recall which is assessed by the open-ended questions. Effective classroom teachers are able to generate accurate diagnoses, not just recognize errors. There appears to be a distinction between knowledge of diagnostic skills and the application of that knowledge.

Teacher performance on the Diagnostic Film Test was analytically related to teacher performance on other reading and reading-related tests, to observed teacher responses to pupils in their classrooms, and to observed and measured pupil reading performances. Some of the results were:

1. *General Findings*:

    a. Second-grade teachers were more skilled at diagnosing problems and suggesting strategies for the remediation of pupil *decoding* problems in oral reading than were fifth-grade teachers.

    b. Second and fifth-grade teachers were about equally skilled at diagnosing problems and suggesting strategies for the remediation of pupil *comprehension* problems in oral reading.

    c. Neither second nor fifth-grade teachers did particularly well in generating diagnoses and remediation strategies using the comparative standard of possible scores.

2. *Performance on Other Reading and Reading-Related Tests:*

    a. The diagnostic and remediation skills of fifth-grade teachers were related to their general knowledge about the teaching of reading.

    b. Teacher skills in the diagnosis and remediation of decoding problems were associated with teacher knowledge about the teaching of mathematics at both grade levels.

3. Observed Teacher Responses to Pupils During Reading Instruction:

a. Second-grade teachers who were more frequently observed helping individual pupils during reading were the more skilled diagnosticians.

b. More highly developed diagnostic skills in second-grade teachers were observationally associated with less of a tendency not to respond to individual pupils during reading activities.

4. Observed Pupil Performances in Reading:

a. Instances of pupil skills in phonics were observed more frequently in the classrooms of diagnostically skilled second-grade teachers, but the same teacher skills were associated with poorer pupil comprehension skills.

b. Instances of positive pupil comprehension skills and general reading skills were observed more frequently in the classrooms of diagnostically skilled fifth-grade teachers.

5. Measured Pupil Performances in Reading:

a. Second-grade pupil skill-growth in reading achievement and application tasks was negatively associated with teacher diagnostic skills.

b. There was no relationship between pupil skill-growth on decoding tasks and the diagnostic skills of teachers.

All of the results should be regarded as hypotheses to be tested in future research. Teachers who took the test were enthusiastic about the Diagnostic Film Test, which they saw as measuring an important teaching skill in a relevant way. Probably because of the face validity of the film test, the comment, "This is the best test of all those we have taken," was not infrequent. Teachers, as well as the two professional staff members knowledgeable about reading who administered the tests and sought teacher comments, felt that the procedure had potential as both a training and selection tool.

Those kinds of responses are perhaps adequate reasons in themselves for the further development of tests of this sort. But the relationships between the demonstration of teacher skills on the test, the observed evidence of pupil skills in related reading tasks, and the observed patterns of teacher responses to individual pupils all suggest that the Diagnostic Film Test is a test worthy of further research and development.

The relatively poor performance of both second and fifth-grade teachers on the comprehension sections of the Diagnostic Film Test, the poor comprehension task performance of second-grade pupils of diagnostically skilled teachers, and the inverse relationship of teacher skills demonstrated on the test and pupil achievement on literal and inferential comprehension tasks all suggest the need for further exploration of the teaching priorities and skills of elementary school teachers.